

THE ROLE OF TRANSCRIPTION AND CHROMATIN IN MAKING  
LONG-TERM MEMORIES

by

Spencer G. Jones

Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy

at

Dalhousie University  
Halifax, Nova Scotia  
July 2023

Dalhousie University is located in Mi'kma'ki, the  
ancestral and unceded territory of the Mi'kmaq.  
We are all Treaty people.

© Copyright by Spencer G. Jones, 2023

# Table of Contents

List of Tables.....	v
List of Figures.....	vi
Abstract.....	viii
List of Abbreviations and Symbols Used.....	ix
Acknowledgements.....	xi
Chapter 1: Introduction.....	1
1.1 Learning and memory.....	1
1.2 The molecular mechanisms of associative memory formation.....	2
1.3 Chromatin and the SWI/SNF complex.....	4
1.4 <i>Drosophila melanogaster</i> as a model organism for studying memory.....	6
1.5 Conditioned courtship suppression in <i>Drosophila melanogaster</i> .....	7
1.6 Study rationale and primary objectives.....	12
Chapter 2: Methods.....	14
2.1 Fly stocks and genetics.....	14
2.2 Mushroom body specific RNAi knockdowns using the GAL4-UAS system.....	16
2.3 Measuring memory with the courtship conditioning assay.....	17
2.5 Cycloheximide feeding.....	19
2.6 Brain dissections and microscopy.....	19
2.7 Sample collection for transcriptome and chromatin analyses.....	20
2.7.1 Memory time-course collections.....	20
2.7.2 Bap60-KD collections.....	20
2.8 Isolation of nuclei tagged in a specific cell-type (INTACT).....	21
2.9 RNA-sequencing.....	22
2.10 ATAC-sequencing.....	22
2.11 Bioinformatic analysis of sequencing data.....	23

2.11.1 RNA-sequencing analysis .....	23
2.11.2 ATAC-seq analysis .....	28
2.12 ChIP-seq analysis .....	31
2.13 MB-specific Brm-TaDa .....	32
 Chapter 3: The transcription factor hierarchy activated during long-term memory formation ....	 33
3.1 Validating established memory formation dogma in the long-term courtship suppression memory assay .....	 33
3.1.1 Translation is required for long-term courtship memory but is dispensable during training and short-term memory formation .....	 34
3.1.2 <i>CrebB</i> is required in the MB for STM and LTM formation.....	37
3.2 INTACT obtains MB-specific nuclear RNA for profiling a time-course of LTM transcription .....	 40
3.3 Identifying transcriptionally regulated genes during LTM formation.....	42
3.4 MB-specific transcript induction occurs as memory training progresses.....	44
3.5 Metabolism and memory signaling are key components of the MB-specific memory transcriptional response .....	 46
3.6 <i>Hr38</i> and <i>sr</i> are candidate courtship memory IEGs induced during long-term courtship memory .....	 48
3.7 <i>Hr38</i> and <i>sr</i> are required in the MB for LTM, but not STM .....	52
3.8 <i>Hr38</i> and <i>sr</i> are highly accessible in the MB at regions of <i>CrebB</i> binding.....	55
3.9 Identifying and characterizing binding sites for <i>Hr38</i> , <i>sr</i> and <i>CrebB</i> .....	57
3.10 <i>sr</i> -bound TIGs show a late MB-specific response .....	59
 Chapter 4: SWI/SNF-mediated chromatin accessibility changes of memory-associated genes during long-term memory formation .....	 61
4.1 LTM is abolished following adult-specific knockdown of <i>Bap60</i> in the MB .....	61
4.2 <i>Bap60</i> -KD suppresses transcription and chromatin accessibility post-training .....	64
4.3 <i>Bap60</i> -dependent memory genes play integral roles in LTM formation.....	69
4.4 <i>Bap60</i> regulates general transcript inducibility of memory-relevant genes during LTM.....	 73

4.5 Bap60-KD does not impact transcription of <i>Hr38</i> and <i>sr</i> in the MB .....	75
4.6 SWI/SNF and <i>sr</i> bind to a subset of memory-relevant Bap60-dependent genes.....	78
Chapter 5: Discussion .....	83
5.1 The transcriptional trace of memory becomes activated in the MB towards the end of training.....	83
5.2 <i>Hr38</i> and <i>sr</i> play a critical role as IEGs in <i>Drosophila</i> courtship memory .....	85
5.3 The SWI/SNF complex is required for memory gene inducibility in the MB during LTM formation.....	87
5.4 Research limitations .....	88
5.5 Conclusions, future directions, and research implications .....	90
References.....	94
Appendix A: List of fly stocks used in this project.....	109
Appendix B: Profiling chromatin accessibility in INTACT-isolated MB nuclei.....	111
Appendix C: Code for processing RNA-seq data.....	112
Appendix D: Code for processing ATAC-seq data .....	113
Appendix E: MB or WH enriched genes identified using INTACT in adult flies.....	114
Appendix F: TIGs identified in both the WH and MB, WH-only or MB-only .....	121

## List of Tables

Table 2.1 List of control and knockdown genotypes used for experiments in this study .....	15
Table 2.2 Read distribution for RNA-sequencing of INTACT-isolated MB samples .....	25
Table 2.3 Read distribution for RNA-sequencing of INTACT-input WH samples .....	26
Table 2.4 Read distribution for RNA-sequencing of Bap60-KD & mCherry-KD MB samples .....	27
Table 2.5 Read distribution for ATAC-sequencing of Bap60-KD & mCherry-KD MB samples .....	30
Table 2.6 Transcription factor ChIP-seq data obtained from ENCODE.....	31
Table 4.1 Bap60-dependent memory genes and presence of SWI/SNF and sr binding sites .....	80

## List of Figures

Figure 1.1 General schematic of courtship memory assay .....	11
Figure 3.1 Cycloheximide-mediated translation inhibition prevents LTM, but not STM formation.....	36
Figure 3.2 MB-specific knockdown of <i>CrebB</i> causes defects in STM and LTM.....	39
Figure 3.3 Visualization of MB and WH-enriched differentially expressed genes and overlap with previously published tissue-specific genes .....	41
Figure 3.4 Differential expression analysis between trained and naive flies during a time-course of LTM formation .....	43
Figure 3.5 Training induced gene transcription becomes more divergent as courtship training progresses.....	45
Figure 3.6 Gene ontology analysis of training induced genes in different tissue-subsets during LTM formation.....	47
Figure 3.7 Transcription of known <i>Drosophila</i> and fly orthologs of mammalian immediate early genes during LTM formation .....	51
Figure 3.8 Adult-specific mushroom body knockdowns of IEGs <i>Hr38</i> and <i>sr</i> in the MB causes defects in LTM but not STM .....	53
Figure 3.9 <i>CrebB</i> binds to accessible chromatin regions of <i>Hr38</i> and <i>sr</i> in the MB.....	56
Figure 3.10 Annotation and characterization of <i>Hr38</i> , <i>CrebB</i> and <i>sr</i> memory transcription factor binding sites.....	58
Figure 3.11 Temporal analysis of MB training induced genes bound by courtship memory transcription factors during LTM formation .....	60
Figure 4.1 Adult-specific knockdown of <i>Bap60</i> in the MB causes defects in LTM, but not STM .....	63
Figure 4.2 Approach for profiling the gene regulatory role of SWI/SNF during LTM formation.....	66
Figure 4.3 Identifying <i>Bap60</i> -dependent memory-regulated transcript level changes.....	67
Figure 4.4 Identifying <i>Bap60</i> -dependent memory-regulated chromatin accessibility changes....	68
Figure 4.5 <i>Bap60</i> regulates both the transcription and chromatin accessibility increases of a subset of memory-relevant genes.....	71
Figure 4.6 <i>Bap60</i> directly and indirectly regulates gene inducibility during LTM formation.....	74

Figure 4.7 Bap60 regulates memory-gene transcription downstream of Hr38 and sr .....	77
Figure 4.8 Identifying and characterizing SWI/SNF and sr-binding sites for Bap60-dependent memory genes .....	79
Figure 4.9 SWI/SNF and sr have proximal or overlapping binding sites for a subset of Bap60-dependent memory genes .....	82
Figure 5.1 Developing a model of courtship memory transcription factor hierarchy .....	93

## Abstract

Long-term memory (LTM) requires transcription and accessible chromatin for formation, however, the genes that are activated and how they promote LTM remains unclear. Here, we profiled the memory transcriptome induced in memory-associated mushroom body (MB) nuclei of *Drosophila melanogaster* at eight time-points during and after courtship memory training. We identify a transcriptional program that becomes activated in the MB as training progresses, enriched for genes in metabolism and known memory signaling pathways, and is distinct from whole-head tissue. Among MB training induced genes, we identified two known insect activity-regulated genes, *Hr38* and *stripe (sr)*. We tested *Hr38* and *sr* for a role in courtship memory following adult-specific knockdown and found that they are required in the MB specifically for LTM formation, with no impact on short-term memory. Further, we show that *Hr38* and *sr* have highly accessible chromatin in the MB at CrebB binding sites, suggesting they are downstream targets of CrebB during LTM formation. To identify what genes could be regulated downstream of *Hr38* and *sr*, we used publicly available binding site information and contrasted this with our memory transcriptome. We find that *sr* may regulate the later MB-specific transcriptional program that occurs when memory consolidation is thought to occur. Our research demonstrates that the SWI/SNF chromatin remodelling complex is required for LTM-associated gene expression and chromatin accessibility changes that occur following memory training, and that this occurs downstream of *Hr38* and *sr*. Finally, we show that disrupting the SWI/SNF complex during LTM results in a near-complete loss of inducible expression of genes with known memory-related functions, highlighting the direct and indirect role SWI/SNF plays in establishing LTM. Collectively, this work provides an invaluable resource of transcription and chromatin regulation during LTM formation for the scientific community, and together with the first mechanistic information about the SWI/SNF complex during acute memory processes, has important implications for understanding processes crucial to cognition in health and disease.

## List of Abbreviations and Symbols Used

<i>Act5c</i>	<i>actic-5C</i>
AMPAR	$\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor
AP-1	Activator protein-1
ARG	Activity regulated gene
AT	After training
ATAC	Assay of transposase accessible chromatin
ATP	Adenosine triphosphate
BAF	BRG1-associated factor
BAP	Brahma associated protein
<i>Bap60</i>	<i>Brahma-associated protein 60 kiloDaltons</i>
BDSC	Bloomington <i>Drosophila</i> Stock Center
bp	Base-pair
<i>brm</i>	<i>Brahma</i>
Ca <sup>2+</sup>	Calcium <sup>2+</sup>
cAMP	Cyclic adenosine monophosphate
CBP	cAMP binding protein
ChIP	Chromatin immunoprecipitation
CI	Courtship index
CREB	cAMP response element-binding protein
<i>CrebB</i>	Cyclic-AMP response element binding protein B
CRTC	CREB-regulated transcription coactivator
CVA	11-cis-Vaccenyl acetate
CXM	Cycloheximide
DE	Differential expression
DEG	Differentially expressed gene
DNA	Deoxyribonucleic acid
DT	During training
ENCODE	Encyclopedia of DNA elements
FDR	False discovery rate
GAL4	Galactose responsive transcription factor GAL4
GAL80 <sup>ts</sup>	Temperature sensitive GAL80
GFP	Green fluorescent protein
GO	Gene ontology
GPCR	G protein-coupled receptors
GTP	Guanosine triphosphate
hpRNA	Hairpin RNA
<i>Hr38</i>	<i>Hormone receptor-like 38</i>
IEG	Immediate early gene
INTACT	Isolation of Nuclei Tagged in a Cell Type

KC	Kenyon cell
KD	Knockdown
LTM	Long-term memory
MB	Mushroom body
mCD8	Mouse cluster of differentiation 8
MI	Memory index
N	Naive
NDD	Neurodevelopmental disorder
NMDAR	N-methyl-D-aspartate receptor
PBAP	Polybromo associated protein
PDE	Phosphodiesterase
PKA	Protein kinase A
PMF	Premated female
<i>pros</i>	<i>Prospero</i>
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
<i>rut</i>	<i>Rutabaga</i>
<i>sr</i>	<i>stripe</i>
SRG	Secondary response gene
STM	Short-term memory
<i>sty</i>	<i>Sprouty</i>
SWI/SNF	SWItch/Sucrose Non-Fermenting
T	Trained
TaDa	Targeted DamID
TES	Transcription end site
TIG	Training induced gene
TRG	Training repressed gene
TRiP	Transgenic RNAi project
TSS	Transcription start site
<i>tutl</i>	<i>Turtle</i>
UAS	Upstream activating sequence
<i>unc84</i>	<i>uncoordinated 84</i>
WH	Whole head

## Acknowledgements

Before you begin reading my thesis, there are a few people that need to be specifically addressed and properly thanked for their contributions and support in making this project and this work possible.

I would like to express my deepest gratitude to my supervisor, Dr. Jamie Kramer, for his unwavering guidance, patience, and invaluable support throughout my research journey thus far. Eight years is a long time to work together, and I am thankful for your expertise, mentorship, and dedication to my academic growth. Since Day 1, the Kramer Lab has been a place where any scientific curiosity could be nurtured into fully grown projects. That is a true testament to the research environment that embodies the essence of science, and I hope that it never changes.

Throughout my journey in the Kramer Lab, I have had the pleasure of working with a wide variety of individuals that have made the past eight years special. This includes Melissa Chubak, Taryn Jakub, Crystal Keung, Taylor Lyons, and Max Stone. An extra big thank you to ‘The Boys’ – Kevin Nixon and Nicholas Raun. The invaluable scientific discussions, critical tech support, and companionship have helped shape this work and made this journey worthwhile.

I would like to express my heartfelt appreciation to my advisory committee, comprising Dr. Barbara Karten, Dr. Deniz Top and Dr. Ian Weaver, for their valuable insights, constructive criticism, and continuous support throughout my time at Dalhousie. Their expertise in their respective fields has greatly enriched the scientific quality of this thesis.

Finally, I am profoundly grateful to my parents, Nancy and Ray Jones, my grandparents Doreen and Bruce Gordon, and my girlfriend Annette Amenta, for their unwavering love, encouragement, and belief in my abilities. To complete a PhD is an extremely privileged position to be in, and their constant support has allowed me to overcome challenges and reach this significant milestone in my life.

*~ In the little grey cells of the brain lies the solution of every mystery ~*

Hercule Poirot (*The King of Clubs*, Agatha Christie)

## Chapter 1: Introduction

Learning and memory, broadly defined, encompasses the encoding, storage, and retrieval of information, resulting in lasting modifications to behaviour in response to changes in the environment (Sweatt, 2010). While humans have had an enduring curiosity about memory since antiquity, we have only recently developed the appropriate tools to study the underlying mechanisms of memory. To date, there is still much to unravel regarding the cellular and molecular changes required to convert transient environmental stimuli into long-lasting behavioural changes. In this thesis, I provide novel insight into the transcriptional and epigenetic mechanisms underlying long-term memory (LTM) formation in memory-specific tissue of the model organism *Drosophila melanogaster*.

### 1.1 Learning and memory

In general, learning and memory processes can be temporally divided into three distinct phases: acquisition or learning, short-term memory (STM) and long-term memory (LTM) (Hawkins et al., 2006; Tully et al., 2003). Learning is the perception of a new experience and serves as the starting point for the acquisition of new information and can be accomplished through either non-associative or associative processes (Lau et al., 2013). Non-associative forms of learning include sensitization and habituation and involve modifications in an organism's initial response to a stimulus following repeated exposure, with sensitization leading to an amplified response and habituation resulting in a diminished response (Groves & Thompson, 1970; Pinsker et al., 1970). Conversely, associative learning occurs when an organism forms connections between two external stimuli, linking one stimulus with another to elicit a specific behavioural response in anticipation of an associated outcome. For example, in associative fear learning, when rodents experience an aversive shock, they exhibit freezing behaviour. Through repeated associations of the shock with a neutral tone or light, the rodents eventually start freezing in response to the tone or light alone (Domjan, 2005).

For an organism to demonstrate meaningful behavioural changes resulting from learning, it is necessary to store the acquired knowledge from training and subsequently retrieve it. This process of storing and retrieving information is commonly referred to as memory. STM may be formed following a brief training period, is malleable and transient, and is reflected by a rapid

decay in the newly learned response. Under appropriate conditions, through sufficient and repeated training, this experience can undergo consolidation, leading to the formation of a persistent long-term memory (LTM) (Quinn et al., 1974). Importantly, the mechanisms underlying associative STM and LTM are highly conserved across animal species. This underscores the importance of learning and memory processes for environmental adaptation and has positioned associative memory as one of the most extensively studied modalities of memory (Milner et al., 1998).

## **1.2 The molecular mechanisms of associative memory formation**

Research pioneered in the sea slug *Aplysia californica* and the fly *Drosophila melanogaster* has provided insight into the signaling pathways required for initiating the formation of associative STM and LTM (Brunelli et al., 1976; Dudai et al., 1976; Hawkins et al., 2006). One of the most well-characterized memory signaling pathways is the cyclic adenosine monophosphate (cAMP) pathway, which has been shown to be essential to form memory in virtually all animal species (Squire, 2009). Focusing on what we have learned about cAMP signalling using the humble fly, in general, cAMP signal transduction begins with ligand binding to G-protein coupled receptors (GPCRs). Ligand binding to GPCRs releases a GTP bound  $\alpha$  subunit, which in turn triggers the activation of adenylyl cyclase, which catalyzes the formation of cAMP from ATP (Bolduc & Tully, 2009; Connolly et al., 1996; Levin et al., 1992; Livingstone et al., 1984; Schwaerzel et al., 2003). In addition to GPCR activation, adenylyl cyclases are also responsive to  $\text{Ca}^{2+}$  signaling. When calcium-permeable glutamate receptors are activated, such as NMDA and AMPA (Davis, 2011), it induces an influx of  $\text{Ca}^{2+}$  into the neuron. High levels of intracellular  $\text{Ca}^{2+}$  then bind to calmodulin, which subsequently activates adenylyl cyclases, leading to an increase in cAMP synthesis. The integration of signals acting upon two molecular pathways highlights the role of adenylyl cyclases as coincident detectors, coordinating cellular responses to both external and internal stimuli.

Synthesis of cAMP is antagonized by cAMP-specific phosphodiesterase's (PDE), which catalyze the degradation of cAMP into adenosine monophosphate (AMP) (Dudai et al., 1976). With repeated concurrent pairing of environmental stimuli, enough adenylyl cyclase becomes sufficiently stimulated to gradually overcome the activity of PDE, resulting in increased levels of intracellular cAMP that can initiate cellular changes. Downstream, cAMP acts to facilitate the activation of cAMP-responsive protein kinases, like Protein Kinase A (PKA). PKA is a tetramer composed of two regulatory and two catalytic subunits. High concentrations of cAMP release PKA

regulatory subunits from the catalytic subunits, enabling the kinase activity of PKA to phosphorylate downstream elements responsible for both STM and LTM (Drain et al., 1991). In STM, short bursts of PKA activity can inhibit S-type K<sup>+</sup> channels, temporarily increasing cellular excitability and synaptic strength (Drain et al., 1991; Goodwin et al., 1997). Extended and repeated bursts of PKA activity is essential to create robust LTM that persists beyond STM generated from single or massed training sessions. Specifically, sustained activity of PKA facilitates its nuclear translocation where it phosphorylates additional targets required for LTM formation. One of the most studied downstream effectors of PKA in the nucleus is the transcription factor cAMP-response element binding protein (CREB) (Tully et al., 1994). CREB binds to cAMP- response elements (CRE) in the genome where it acts to regulate the transcription of genes, often in combination with various recruited cofactors including cAMP binding protein (CBP) and CREB-regulated transcription coactivator (CRTC) (Hirano et al., 2016). Notably, and of critical importance to this work, is that it has been consistently demonstrated across species that transcription, regulated in part by CREB, and subsequent *de novo* protein synthesis is required in the formation of LTM, while these processes are not necessary for STM (Montarolo et al., 1986; Montminy et al., 1986; Smolik et al., 1992; Tully et al., 2003; Yin et al., 1994).

While the signaling pathways that initiate transcription during LTM are relatively well understood, we only have a loose understanding of what occurs next (Gil-Marti et al., 2022; Yap & Greenberg, 2018). What is generally accepted to occur is that during memory training, constitutively expressed transcription factors, including CREB, are activated, which triggers a wave of immediate early gene (IEG) transcription. These IEGs in turn regulate the expression of downstream secondary response genes (SRGs) (Giorgi & Marinelli, 2021). Until recently it was thought that there was a common set of IEGs that are almost always strongly induced in neurons in response to memory training. However, it is now well accepted that both cell-type and memory paradigm contribute to a significantly diverse IEG transcriptional response during LTM formation, although the core primarily consists of established IEGs (Chen et al., 2016; Gil-Marti et al., 2022; Hrvatin et al., 2018; Wu et al., 2017). Core IEGs include the well characterized mammalian IEGs *c-fos*, *arc*, *egr1*, and *Npas4* (Kaldun & Sprecher, 2019). The precise timing of transcription and protein synthesis within memory specific brain regions is crucial for the establishment of an enduring memory. To better understand the mechanisms underlying IEG and SRG transcription

and how this promotes the formation of a stable LTM this work profiles the memory-transcriptome starting from the beginning of memory training until when a LTM can be recalled (Chapter 3).

### **1.3 Chromatin and the SWI/SNF complex**

Broadly defined, epigenetics refers to the study of gene expression changes that occur without alterations to the underlying DNA sequence (Neigeborn & Carlson, 1984). In eukaryotes, DNA is organized and tightly compacted into highly ordered chromatin structures. The fundamental structural unit of chromatin is the nucleosome. Nucleosomes consist of approximately 147 base pairs of DNA wrapped around a histone octamer (Luger et al., 1997). Apart from allowing long stretches of DNA to fit into the cell nucleus, nucleosomes play a critical role in regulating gene expression. Specifically, nucleosome positioning is dynamic and how compact nucleosomes are to each other is crucial in determining the accessibility of DNA to the transcriptional machinery required for gene activation (Schones et al., 2008). Nucleosome positioning is directed by various biochemical modifications that promote chromatin condensation or relaxation, ultimately leading to the repression or activation of gene expression. These modifications can include DNA methylation, histone post-translational modifications, and ATP-dependent chromatin remodeling (Goldberg et al., 2007).

The SWI/SNF (SWItch/Sucrose Non-Fermenting) complex is an ATP-dependent chromatin remodeling complex (Neigeborn & Carlson, 1984; Stern et al., 1984). First described in yeast, the complex has since been shown to be highly conserved across species, including humans, flies, and rodents (Son & Crabtree, 2014). The SWI/SNF complex plays a critical role in regulating gene expression by modifying the structure of chromatin. In general, the SWI/SNF complex is recruited to regions of active chromatin by binding to the activation domains of DNA-bound transcription factors (Clapier & Cairns, 2009; Yudkovsky et al., 1999). Then, by harnessing the energy generated from ATP hydrolysis, the SWI/SNF complex destabilizes the DNA-histone interaction, generating loops in the DNA that facilitate the movement of nucleosomes along the DNA molecule (López & Wood, 2015; Vogel-Ciernia et al., 2013a). As a result of this shift in nucleosome position, DNA becomes more accessible to transcriptional machinery, leading to increased gene transcription (Biggar, 1999; Kassabov et al., 2003; Lomvardas & Thanos, 2001; Lorch et al., 2001).

In mammals, the SWI/SNF complex is known as the BAF complex and consists of 15 subunits encoded by 29 genes from 15 gene families (Son & Crabtree, 2014). To date, 15 of 29 genes have been implicated in intellectual disability, which has put an emphasis on understanding the role that the SWI/SNF complex plays in neurons (Jakub et al., 2021). What has been discovered thus far is that functionally, the BAF complex plays a critical role in neurodevelopment, particularly during neuronal differentiation and establishing neuronal identity (Hargreaves & Crabtree, 2011; Lessard et al., 2007; Narayanan et al., 2015; Narayanan & Tuoc, 2014; Olave et al., 2002; Tuoc et al., 2017). Loss of the core BAF subunits BAF170 and BAF155 in mouse models results in disrupted neurodevelopment, behavioral deficits, and impaired learning (Tuoc et al., 2017). Further, loss of these subunits is associated with a decrease in active chromatin marks, and an increase in repressive marks (Narayanan et al., 2015). This suggests that SWI/SNF is required for gene transcription activation during neurodevelopment by interacting with histone modifying enzymes. The nBAF complex, which consists of a distinct composition of BAF subunits in neurons, has also been shown to play a crucial role in adult cognition (López & Wood, 2015; Tuoc et al., 2017; Vogel-Ciernia et al., 2013a; Wu et al., 2007). Mice that have been genetically modified to express a mutant nBAF complex in post-mitotic neurons have displayed deficiencies in synaptic plasticity and impaired memory formation (Vogel-Ciernia et al., 2013a). Interestingly, recent evidence in mice fibroblasts suggests that AP-1, an IEG transcription factor complex, can actively recruit the BAF complex to enhancers to establish accessible chromatin in response to environmental stimuli (Vierbuchen et al., 2017). Currently, no investigations have examined whether neurons exhibit a parallel mechanism of BAF-mediated regulation of activity-dependent gene programs. Nonetheless, the potential for a such a mechanism, particularly within the realm of memory due to the BAF complex's established role, remains a promising avenue for exploration.

In *Drosophila*, the SWI/SNF complex exists in two distinct forms: the BAP and PBAP complexes (Mashtalir et al., 2018). Functionally, the SWI/SNF complex in *Drosophila* exhibits diverse roles in neuronal development and function, including involvement in dendrite arborization and targeting, neuron remodeling and memory formation (Chubak et al., 2019; Kirilly et al., 2011; Nixon et al., 2019; Parrish et al., 2006; Tea & Luo, 2011). Work in our lab has highlighted the importance of the SWI/SNF complex in mushroom body (MB) memory neurons. In an RNAi screen, different SWI/SNF subunits were knocked down in MB neurons. Core

SWI/SNF subunits Bap60 and Snr1, along with the PBAP-specific subunit E(y)3, were found to have deficiencies in MB neuron remodeling (Chubak et al., 2019). We also identified a role for the SWI/SNF complex in MB for both STM and LTM. For some SWI/SNF components, loss of STM correlated with defects in MB remodeling; however, several subunit knockdowns resulted in memory defects that occurred independent of remodeling defects (Chubak et al., 2019). Further, knockdown immediately following eclosion of the core SWI/SNF subunit, Bap60, which correlates with early juvenile development, resulted in a LTM defect (Nixon et al., 2019). Collectively, these findings suggest that the SWI/SNF complex plays a distinct role during neurodevelopment in memory formation within the *Drosophila* brain, which aligns with the evidence from mammals. This study explores the SWI/SNF complex's unexplored role in acute memory processes, providing valuable insight into its functions in neurons, independent of development (Chapter 4).

#### **1.4 *Drosophila melanogaster* as a model organism for studying memory**

*Drosophila melanogaster*, commonly referred to as the fruit fly, has played a critical role in genetic research since the first white eyed mutant flies (as opposed to the natural red eye) were identified by T.H. Morgan at the beginning of the 20<sup>th</sup> century (Morgan, 1910). Since then, the fly, aided by its ease of culture, accessible genetic tools, and high degree of genetic homology to human disease genes, has become a well-established model in the study of brain function and memory (Coll-Tane et al., 2019; Roote & Prokop, 2013; Tian et al., 2017). In fact, mutagenesis screens using aversive olfactory memory in flies were the first to identify single-gene memory mutants, including *amn*, *dnc*, and *rut* (Dudai et al., 1976; Livingstone et al., 1984; Siegel & Hall, 1979). These memory mutants all encode components of the cAMP pathway and paved the way for making it one of the most well-studied molecular mechanisms underlying memory (Brunelli et al., 1976).

The centre for olfactory memory in the fly brain is the mushroom body (MB), a structure which at times has been argued to be analogous to mammalian brain structures, including the hippocampus—a key brain centre involved in learning and memory (Coll-Tane et al., 2019; Han et al., 1992; Heisenberg et al., 1985). Chemical ablation of the MB impairs both STM and LTM across various *Drosophila* learning and memory paradigms, while not causing immediate lethality, highlighting the importance of this structure for memory (de Belle & Heisenberg, 1994; McBride

et al., 1999). Structurally, the MB appears as a pair of synaptically dense neuropils consisting of ~2200 neurons, called Kenyon cells (KC), with three distinct neuronal subtypes ( $\alpha/\beta$ ,  $\alpha'/\beta'$ , and  $\gamma$ ) that contribute to the formation of five distinct lobes (Aso et al., 2014). The different lobes of the MB are responsible for processing different phases of memory, with the  $\alpha/\beta$ ,  $\alpha'/\beta'$  lobes primarily processing LTM, and the  $\gamma$  lobe primarily processing STM (Krashes et al., 2007; Montague & Baker, 2016; Trannoy et al., 2011). On a molecular level, many components of the cAMP signaling pathway are highly expressed in the MB including PDE (encoded by *dnc*), AC (*rut*) and *CrebB* (Blum et al., 2009; Dudai et al., 1976; Livingstone et al., 1984; Zhang et al., 2015). Taken together, the MB is the logical area to investigate the transcriptional trace of memory that is formed during LTM.

### **1.5 Conditioned courtship suppression in *Drosophila melanogaster***

Conditioned courtship suppression, also known as “courtship conditioning”, and hereafter referred to as courtship memory, is a learning and memory assay which takes advantage of naturally occurring, hard-wired, male courtship behaviour (Raun et al., 2021; Siegel & Hall, 1979; Spieth, 1974). Courtship behaviour refers to the set of highly stereotyped behaviours that male fruit flies display upon being exposed to a potential mate. These distinct behaviours include orienting towards the female, tapping her with his forelimb, contacting her genitals using his proboscis and generating a courting song by vibrating an outstretched wing. The culmination of these behaviors is an attempt at copulation, to which the female can respond either positively by spreading her wings to indicate willingness, or negatively by kicking to repel the male (Koemans, Oppitz, et al., 2017; Spieth, 1974).

While male courtship behaviour is hard-wired, with the underlying neural circuitry likely to be fully mapped in the coming years (Clowney et al., 2015; Pavlou & Goodwin, 2013), male flies can learn to adapt their courtship behavior over time, selectively directing it towards individuals more likely to engage in mating. Adaptive male courting behaviour forms the basis of the courtship memory assay. In the courtship memory assay, a newly eclosed, socially naive male is paired with a single recently-mated female, which actively reject further courtship attempts. During the courtship training period, the male fly actively attempts to court and mate with the female but is rejected. As a result of the failed mating attempts, the male fly decreases its courtship behaviour. This trained courtship suppression can be recalled upon during subsequent encounters

with a female fly. It is thought that the courtship memory formed by the male fly during training is due to learning to associate the physical rejection of the failed copulation attempts with the olfactory profile of the female, which includes the male pheromone cis-vaccenyl acetate (cVA) that is deposited on the cuticle of the female by the male during mating (Ejima et al., 2007; Keleman et al., 2012). Both STM and LTM can be formed using the courtship memory assay by simply altering the length of the courtship training period (**Figure 1.1**). A one to two-hour training can generate a STM which can be recalled up to three hours later. Conversely, a single five to seven-hour courtship training session can form a persistent LTM which can be recalled up to a week later (Keleman et al., 2007, 2012; McBride et al., 1999). Taken together, the courtship memory assay is a valuable tool to study memory processes and offers a robust, more ethologically relevant form of memory compared to the commonly used aversive olfactory memory assays that involve pairing an electrical shock with an odor.

There are two types of courtship memory assays described in the literature: cVA-retrievable and associative courtship memory (**Figure 1.1**) (Raun et al., 2021). Methodologically, these two courtship memory assays are distinguished by the way in which the formed courtship memory is recalled. In cVA-retrievable courtship memory, mated females are used during training, as well as in memory testing. Here, memory recall is based on the presence of cVA. While cVA has a natural anti-aphrodisiac effect, male flies that have previously experienced mating failure show a dramatically enhanced courtship suppression in response to cVA. Conversely, associative courtship memory uses a mated female for training, with memory retrieved using a virgin female. Like recently mated females, newly eclosed virgin females are also unreceptive to mating attempts, and associative courtship memory recall draws upon associations formed during training between mating failure and cuticular pheromones on the female. As both types of courtship memory use the same types of trainer females, it is thought that both courtship memories are formed simultaneously and that differences lie in the neural circuitry used through which the memory is retrieved. Most importantly, however, like other *Drosophila* associative memory assays, courtship memory is dependent on the neural circuitry in the MB for normal STM and LTM formation (McBride et al., 1999).

On a molecular level, the requirement for cAMP signaling is conserved in the formation of associative courtship memory and has been well documented. In fact, the first paper to describe

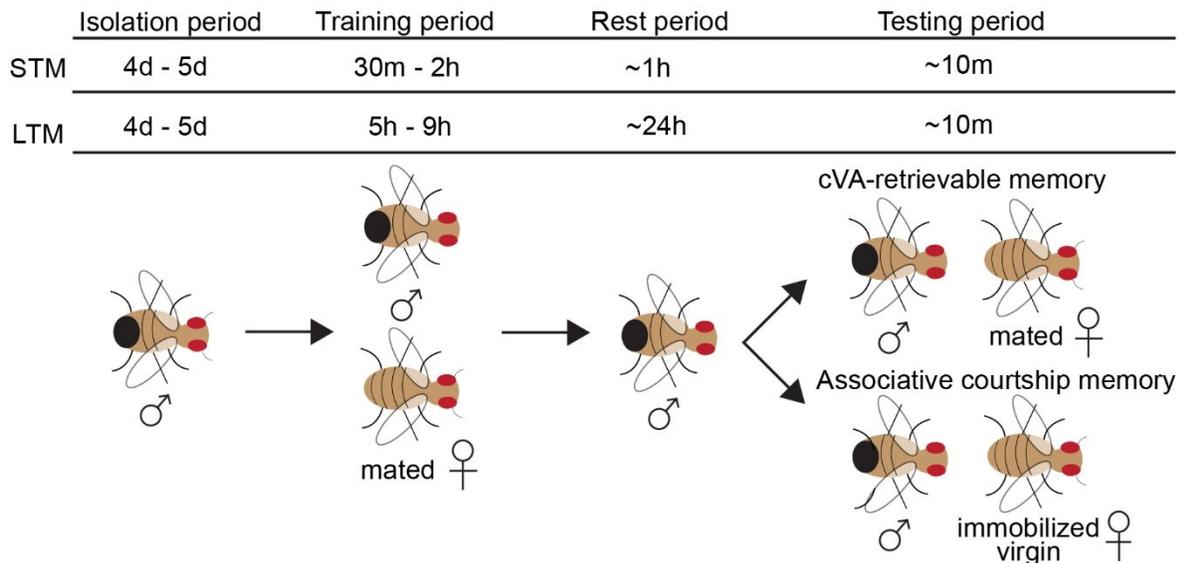
courtship memory identified STM defects in *amn* mutants, which encodes a neuropeptide (Siegel & Hall, 1979). Since then, many of the core components of cAMP signalling have been shown to be required for associative courtship STM, including AC, PDE, PKA and Calmodulin (Ackerman & Siegel, 1986; Broughton et al., 2003; Ejima et al., 2005; Gailey et al., 1984; Joiner & Griffith, 1997; Kane et al., 1997; Mehren & Griffith, 2006; O'Dell et al., 1999; Siegel & Hall, 1979). Several components of the cAMP pathway have also been shown to be required for associative courtship LTM, however, most importantly the requirement for CREB-mediated transcription is conserved, highlighting the similarities of associative courtship memory to other forms of associative LTM (K. Le Li et al., 2018; Sakai et al., 2004).

Associative and cVA-retrievable courtship memory have only recently been distinguished from one another, and many of the cAMP pathway components required for associative courtship memory have not yet been tested for a role in cVA-retrievable courtship memory (Raun et al., 2021). To date, CamKII (calmodulin kinase) is the only component of cAMP signalling which has been explicitly cross-validated between associative and cVA-retrievable courtship memory, with a role in STM formation found in both (Joiner & Griffith, 1997). While it is expected that other major elements of cAMP signalling are conserved between associative and cVA-retrievable courtship memory – including CREB-mediated transcription and downstream translation– recent studies have shown that there are some differences between these two assays. For example, cVA-retrievable courtship memory requires the dopamine GPCR, DopR1, for both STM and LTM, whereas associative courtship memory requires the dual GPCR that is activated by both dopamine and ecdysteroids, DopEcR (Ishimoto et al., 2013; Keleman et al., 2012). This suggests that different cellular and molecular mechanisms underly both cVA-retrievable and associative courtship memory, and neurogenetic results using these different courtship memory recall methods should be interpreted with this in mind.

Courtship memory, which has been highlighted as inducing a robust natural LTM with minimal external interference experimentally, is a unique assay ideally suited to study the transcriptional trace of LTM formation. However, to date, there have only been three studies that have examined the memory-regulated transcriptome formed by courtship memory training (Barajas-Azpeleta et al., 2018; Jones et al., 2018; Winbush et al., 2012). Two of these studies have looked at transcript level differences in the whole head, profiling time points at either one hour or

24 hours after long-term courtship memory training had ended. Together, genes encoding immune peptides with antimicrobial activity were found to be upregulated in the fly head one hour after training, with minimal transcriptional changes at the 24 hours post-training time-point (Barajas-Azpeleta et al., 2018; Winbush et al., 2012). As profiling bulk whole-head tissue can introduce unwanted biological variance which may mask memory-relevant transcriptional changes, current research has shifted towards studying cells which are specifically part of the memory engram – which as previously emphasized, in *Drosophila* are in the MB.

The only study thus far to study transcript level changes in the MB during long-term courtship memory formation was work published by our group (Jones et al., 2018). In this work, we profiled transcript level changes in the MB one hour and 24 hours after long-term courtship memory training had finished. Like the other whole-head courtship memory transcriptome papers, we found that there was more transcriptional activity in the MB immediately following courtship training compared to one-day later, when contrasted with baseline transcript levels in control naive male flies. Among notable training induced genes, we identified the upregulation of transcripts encoding components of the cAMP pathway – some of which have been mentioned – including calmodulin (*Cam*), *DopEcR*, a PKA regulatory subunit (*PKA-R2*), and CBP (*nej*), among others. The results and methods of this work act as a pilot study for portions of this thesis, as much remains to still be discovered from the MB courtship memory transcriptome. Specifically, unlike other *Drosophila* associative memory assays which require manual US/CS spacing to induce LTM, male flies naturally space their mating attempts during courtship memory training. This makes it possible for continuous training session to form a persistent LTM (Keleman et al., 2007, 2012; McBride et al., 1999). While beneficial from an experimental handling stand-point, a continuous LTM training period makes it difficult to ascertain when memory-relevant transcription becomes activated. It is highly likely that profiling the memory transcriptome during the courtship training period will yield interesting and novel insight into the mechanisms of LTM formation. Further, it is possible that there are transcriptionally active time-points beyond one-hour after training, like what has been observed in other associative memory assays, that have yet to be explored in long-term courtship memory formation (Dubnau et al., 2003). Elucidating the temporal gaps of the MB courtship memory transcriptome are a prime focus of this thesis (Chapter 3).



**Figure 1.1 General schematic of courtship memory assay**

The courtship memory assay can be used to induce both a short-term memory (STM) and a long-term memory (LTM). During the isolation period a newly eclosed male fly is individually housed and isolated for 4 to 5 days. During the training period male flies are placed with a single mated trainer female. STM can be induced using a 30 min to two-hour training, while LTM requires 5-9h. Male flies are then separated and re-isolated for 1h (STM) or ~24h (LTM). To test memory recall, male flies are paired with a tester female and courtship behaviour recorded and quantified. Two types of tester females are used – mated tester females (cVA-retrievable memory) or an immobilized virgin tester female (associative courtship memory) – which evoke courtship memory through different mechanisms. Figure adapted from (Raun et al., 2021).

## 1.6 Study rationale and primary objectives

The overall aim of this thesis was, using a cell-type specific genomics approach, to elucidate the transcriptional and epigenetic mechanisms underlying the induction of memory genes during LTM formation. Leveraging the diverse genetic tools available in *Drosophila*, this work looked to capture the full spectrum of memory-sensitive IEGs and SRGs transcribed during LTM formation – which we collectively refer to as training induced genes (TIGs) - in memory-specific tissue of the MB. Specifically, we investigated the MB memory transcriptome during a time-course of LTM formed by the conditioned courtship suppression memory assay beginning during the training period, up until one-day later when LTM is fully established. Initially exploratory, we hypothesized that this work would identify a wide range TIGs, including known neuron-activity regulated IEG transcription factors which would then be tested for a role in LTM formation. We expected that candidate courtship memory IEGs would play an essential role in LTM formation by contributing to the regulation of downstream SRG transcription.

There are many potential epigenetic regulators that could contribute to the dynamic changes in chromatin accessibility required for establishing LTM. The work presented here focuses on exploring the role of a core subunit of the SWI/SNF complex, Bap60, in regulating memory transcription and chromatin accessibility during LTM formation. Previous work in our lab has shown that Bap60 plays an important role in establishing normal chromatin accessibility and transcriptional programs required for proper LTM formation in the *Drosophila* MB during larval and early juvenile development; however, a role for Bap60 in acute memory processes, independent of development, has yet to be determined. Using adult-specific genetic knockdowns in the MB, a role for *Bap60* in memory functioning was determined and using MB cell-type specific transcriptomics and epigenomics the impact of *Bap60* knockdown during LTM formation on memory-regulated transcript levels and chromatin accessibility changes was characterized. Here, our guiding hypothesis was that *Bap60* would be required for forming transcription-dependent LTM, but not STM, and that knockdown of *Bap60* would disrupt normal TIG induction due to a diminished capacity for the SWI/SNF complex to dynamically increase chromatin accessibility during LTM formation.

**Altogether, the specific objectives of this study were to:**

1. Identify temporally regulated TIGs transcribed in the MB during LTM formation
2. Characterize the role of candidate courtship memory IEGs, identified in objective 1, in the MB during LTM formation.
3. Identify and characterize a role for a core subunit of the chromatin regulating SWI/SNF complex, Bap60, in the MB during acute memory processes

This study provides the most extensive look into the temporal dynamics of memory-sensitive gene transcription in memory-specific tissue to date and provides novel insight into the hierarchy of transcription factors that are activated in the MB during LTM formation. Further, this work reveals dynamic training induced chromatin accessibility changes and identifies a novel role for the SWI/SNF chromatin remodelling complex in regulating memory-gene inducibility in the MB during LTM formation. Collectively, this work provides an invaluable resource of transcription and chromatin regulation during LTM formation for the scientific community, and together with the first mechanistic information about the SWI/SNF complex during acute memory processes, has important implications for understanding processes crucial to cognition in health and disease.

## Chapter 2: Methods

### 2.1 Fly stocks and genetics

*Drosophila melanogaster* stocks were reared on a standard medium (cornmeal-sucrose-agar), supplemented with the mould inhibitors methyl paraben and propanoic acid, at 25°C in 70% humidity with a 12-h light/dark cycle. All fly stocks were obtained from the Bloomington *Drosophila* Stock Center (BDSC; Bloomington, USA), with the exception of wild-type female flies which were a Canton-S/Oregon-R mixed genetic background generated by J.M. Kramer, as well as *UAS-Unc84::GFP*, which was a gift from Gilbert L. Henry (Henry et al., 2012). Inducible RNA interference (RNAi) stocks were obtained from the BDSC Transgenic RNAi project (TRiP) library (Perkins et al., 2015). Flies containing the MB-specific *RI4H06-GAL4* were generated by the Rubin lab for the Janelia Research Campus FlyLight project and obtained from BDSC (Jenett et al., 2012). Refer to **Appendix A** for a list of all fly stocks used in this thesis and a brief description.

RNAi lines selected to test candidate genes for a role in memory were identified based on prior phenotype screens or an evaluation of RNAi efficacy using a lethality assay or RT-qPCR (**Appendix A**). For knockdown experiments in courtship conditioning or for sequencing experiments, control crosses were generated in parallel using the appropriate genetic background stocks. Of specific note, in knockdown experiments using the BDSC RNAi line #32503 (**Chapter 3**), which targets a core SWI/SNF subunit Bap60, a hairpin stock targeting the mCherry fluorophore (genotype:  $y^l\ sc^* v^l\ sev^{21}; P\{VALIUM20-mCherry\}attP2$ ) was used as the genetic control. The mCherry fluorophore is a protein that does not exist in *Drosophila*, and controls for the genetic background as well as for the non-specific effects of RNAi (Perkins et al., 2015). A full list of experimental and control genotypes that were the F1 product of crosses used in this thesis for memory testing or sequencing experiments are listed in **Table 2.1**.

**Table 2.1 List of control and knockdown genotypes used for experiments in this study**

*UAS-RNAi* represents generic RNAi stocks for various genotypes. Full stock genotypes for both RNAi and controls are listed in **Appendix A**.

<b>Experiment</b>	<b>Control Genotype</b>	<b>Knockdown Genotype</b>
Courtship conditioning (STM/LTM) – MB knockdown ( <b>Chapter 3</b> )	$\frac{y^l, v^l; R14H06-GAL4; attp40}{Y \quad + \quad +}$	$\frac{y^l, v^l; R14H06-GAL4; UAS-CrebB^{RNAi}}{Y \quad + \quad +}$ <b>Also referred to as <i>CrebB-KD</i></b>
RNA/ATAC-seq ( <b>Chapter 3</b> )	$\frac{\pm; UAS-Unc84::GFP; R14H06-GAL4}{Y \quad + \quad +}$	
Courtship conditioning (STM/LTM) – adult-specific MB knockdown ( <b>Chapter 4</b> )	$\frac{y^l, v^l; tubP-GAL80^{ts}; R14H06-GAL4}{Y \quad + \quad attp2}$	$\frac{y^l, v^l; tubP-GAL80^{ts}; UAS-Hr38^{RNAi}}{Y \quad + \quad attp2}$ <b>Also referred to as <i>Hr38-KD</i></b>
Courtship conditioning (STM/LTM) – adult-specific MB knockdown ( <b>Chapter 4</b> )	$\frac{y^l, v^l; tubP-GAL80^{ts}; R14H06-GAL4}{Y \quad + \quad attp2}$	$\frac{y^l, v^l; tubP-GAL80^{ts}; UAS-sr^{RNAi}}{Y \quad + \quad attp2}$ <b>Also referred to as <i>sr-KD</i></b>
RNA/ATAC-seq and courtship conditioning (STM/LTM) – adult-specific MB knockdown ( <b>Chapter 4</b> )	$\frac{y^l, sc^*, v^l; tubP-GAL80^{ts}; R14H06-GAL4}{Y \quad UAS-Unc84::GFP \quad UAS-mCherry^{RNAi}}$ <b>Also referred to as <i>mCherry-KD</i></b>	$\frac{y^l, sc^*, v^l; tubP-GAL80^{ts}; R14H06-GAL4}{Y \quad UAS-Unc84::GFP \quad UAS-Bap60^{RNAi}}$ <b>Also referred to as <i>Bap60-KD</i></b>

## 2.2 Mushroom body specific RNAi knockdowns using the GAL4-UAS system

MB-specific knockdown of candidate memory genes was achieved by using the GAL4-UAS binary expression system in combination with RNAi mediated knockdown (Brand & Perrimon, 1993). The GAL4-UAS system uses a yeast-derived transcription factor GAL4 to activate the expression of transgenes under the control of the GAL4-specific enhancer UAS (upstream activating sequence). Tissue-specific expression of UAS controlled transgenes can be achieved by expressing GAL4 under the control of one or more transcriptional enhancers. To achieve MB-specific GAL4 expression the transgenic driver *R14H06-GAL4* was leveraged, which expresses GAL4 under control of a MB specific enhancer fragment from the *rutabaga* gene and has been validated to be highly specific to  $\alpha/\beta$  and  $\gamma$  neurons of the MB (Chubak et al., 2019; Jones et al., 2018; Kummeling et al., 2021; Nixon et al., 2019). Of note, while *R14H06-GAL4* was the driver used almost exclusively in this study, for lethality assays KD was induced ubiquitously using an *Act5C-GAL4* driver, which utilizes the *Act5C* promoter.

To generate MB-specific knockdowns (KD), in general, flies carrying a homozygous copy of *R14H06-GAL4* were crossed with flies of the opposing sex homozygous for a *UAS-RNAi* sequence specific to a candidate memory mRNA transcript. In this manner, GAL4 induces the downstream expression of either a short or long hairpin RNA (hpRNA) specifically in MB tissue for RNAi gene knockdown. Mechanistically, hpRNA operates by leveraging the cell's own endogenous proteins for processing small RNA molecules. Specifically, a hpRNA molecule is synthesized or introduced into the cell and is designed to have a region that is complementary to the target gene's mRNA sequence. Once inside the cell, an enzyme called Dicer-2 recognizes and processes the hpRNA into short-double stranded RNA fragments known as small interfering RNAs (siRNA). These siRNAs are then incorporated into RNA-induced silencing complexes (RISCs) which then scan the cell for mRNA molecules that possess sequences complementary to the siRNA. When a target mRNA is encountered, the siRNA guides the RISC complex to bind to the mRNA and initiate degradation.

Two main approaches were used in this work for MB-specific knockdowns. In the first, experimental fly crosses were reared at 25°C, which induces RNAi knockdown during development. This approach was exclusively used for *CrebB*-KD (**Chapter 3.1.2**). For the remainder of knockdown experiments, the temperature-sensitive GAL80 (GAL80<sup>ts</sup>) system was used to induce RNAi knockdown specifically in the adult-fly, which allows for testing the role of

a gene in memory independently from development (Mcguire et al., 2003). GAL80<sup>ts</sup> is a transcription factor that, when expressed, binds to, and inactivates GAL4. By combining MB-specific *R14H06-GAL4* with a ubiquitously expressed GAL80<sup>ts</sup> construct (tubP-GAL80<sup>ts</sup>), a precise temporally controlled knockdown can be achieved. At 18°C, GAL80<sup>ts</sup> is expressed and represses GAL4, whereas at 29°C GAL80<sup>ts</sup> is inactivated, permitting GAL4-mediated transcriptional activation. To achieve adult-specific MB knockdown for genes being tested for a role in memory, flies were raised at 18°C throughout development and for four days post-eclosion. Knockdown was initiated one-day prior to behavioural training by moving flies to 29°C. A 12h-12h light-dark cycle in 70% humidity was used at all temperatures for both the developmental adult-specific knockdown approaches.

### **2.3 Measuring memory with the courtship conditioning assay**

Courtship conditioning was used to test for deficiencies in STM and LTM and was performed as previously described with minor modifications (Koemans, Oppitz, et al., 2017; Siegel & Hall, 1979). In the courtship conditioning assay, a male fly is paired with an unreceptive, previously mated female (PMF). During this training session male flies learn to suppress their courting attempts in response to continual rejection by the PMF. Male flies with normal memory functioning will be able to recall this previous rejection experience and reduce their courting attempts in a subsequent encounter with a different PMF, however, memory deficient flies will not show reduced courting behaviour. Here, we looked to see if flies were memory deficient following RNAi genetic knockdown, as an indicator that the genes being targeted are involved in memory processes.

Described here is the general approach used for courtship conditioning. Newly eclosed F1 male flies were collected and isolated for five days in individual wells of a 96-well block that contained 500µL of media, this is referred to as an isolation chamber. F1 male flies were split into two cohorts – trained and naive flies. Naive flies do not undergo training and act as a control by displaying courting behaviour unaffected by training. On the day of training, five-day old males, both naive and trained, were gently aspirated into individual wells of a freshly made isolation chamber. CO<sub>2</sub> was not used at any point on males beyond eclosion to mitigate any potential extrinsic stress, and naive flies were moved similarly to trained flies to account for behavioural changes that could arise from extra handling. Male flies were trained by introducing an unreceptive

PMF into the isolation chamber. For STM, male flies were trained by pairing with a PMF for two hours and then placed back into isolation for one hour – referred to as the rest period. For LTM, male flies were trained using a single seven-hour training session, followed by re-isolation and a rest period of 20-24h. Following the rest period, courtship activity was measured for each individual naive and trained male fly by pairing with a new PMF in what is referred to as the testing period. For testing, individual male flies (naive or trained) and a new PMF and placed into a custom designed mating chamber that contain eighteen 1 cm diameter mating circles, allowing for 18 fly pairs to be tested simultaneously. Male courtship was recorded using a digital camera for ten minutes.

For every male-female fly pair, a courtship index (CI) was determined by manual visual analysis for the male fly. CI refers to the proportion of time the male fly engaged in courtship behaviour during the ten-minute testing period. Courtship behaviours measured include orienting towards and following the female, tapping her with his forelimb, contacting her genitals using his proboscis and generating a courting song by vibrating an outstretched wing, as well as attempted copulation. Once a CI was obtained for naive and trained cohorts, a memory index (MI) was calculated based on the following formula:  $MI = ((CI_{naive} - CI_{trained}) / CI_{naive})$  and allowed for detection of differences in courtship memory between KD flies and their respective controls (Keleman et al., 2012).

Memory deficiency was statistically tested using two distinct approaches. First, CIs between trained and naive flies of the same genotype were compared to determine if there was a reduction in courtship behaviour. Second, a comparison of MIs between KD and control genotypes was performed to indicate if memory was induced similarly, or specifically, if some memory retention was impaired due to the KD. To prepare CI data for analysis, outliers were first removed by GraphPad Prism (v 9.5.1) using the ROUT method with the false discovery rate set at the default value of 1%. Statistical analysis of CIs was then performed between naive and trained flies using a two-tailed Mann-Whitney test. If  $P < 0.05$  between naive and trained CI (when mean naive CI > mean trained CI), then this is an indication that the genotype has retained the memory formed during training and male flies were able to reduce their courting behaviour during the testing period. For statistical comparison between the MIs of KD and control genotypes a randomization test (random sampling with replacement, 10000 replicates) was performed using a custom bootstrapping R script (Koemans, Kleefstra, et al., 2017). A significant reduction in MI ( $P < 0.05$ )

indicates that some level of memory retention was lost due to the KD, whereas  $P > 0.05$  is an indication that normal memory retention between KD and control was observed. Visualization of CI and MI data was generated using GraphPad Prism, with CI values presented as box and whisker plots, with whiskers showing values in the 5-95<sup>th</sup> percentiles.

## 2.5 Cycloheximide feeding

To determine if translation was required during courtship memory formation, cycloheximide was fed to wild-type male flies (Canton-S/Oregon-R) to block protein synthesis. Courtship isolation chambers were filled with media consisting of either 1% agarose, 5% sucrose and 35 mM cycloheximide dissolved in 3% ethanol (sucrose + CXM), or just sucrose and agarose (sucrose-only) as a control, as previously described (Tully et al., 1994). Flies were raised on standard media and transferred to sucrose + CXM or sucrose-only isolation chambers one day prior to STM or LTM training. For STM, flies rested on sucrose + CXM prior to testing. For LTM, flies either rested on sucrose + CXM or returned to standard media prior to testing.

## 2.6 Brain dissections and microscopy

To look for gross morphological defects in the development of the MB following *CrebB*-KD, male adult fly brains were dissected from the F1 generation of the cross between *UAS-mCD8::GFP*, *R14H06-GAL4* and either *CrebB*-KD or the corresponding control background. *UAS-mCD8::GFP* produces a cell surface glycoprotein fused to GFP, which when driven by *R14H06-GAL4* allows for clear MB visualization. Brains were dissected in PBS and fixed with 4% paraformaldehyde for 45 minutes at room temperature. Brains were then imaged using a Zeiss LSM 710 confocal fluorescent microscope, processed using ImageJ (v 1.53K) and the MB manually scanned for defects, in an approach previously described, which potential defects including: the appearance of missing  $\alpha$  or  $\beta$  lobes,  $\beta$ -lobe fibers crossing over the midline, extra dorsal projects and/or stunted  $\gamma$ -lobes (Chubak et al., 2019). Dissections and confocal microscopy were performed by Olivia Kerr.

## 2.7 Sample collection for transcriptome and chromatin analyses

### 2.7.1 Memory time-course collections

Male flies homozygous for *UAS-Unc84::GFP ; R14H06-GAL4* were crossed to virgin females of a Canton-S/Oregon-R genetic background. *UAS-Unc84::GFP* is a protein expressed that allows for the immunoprecipitation of MB neuron nuclei downstream (Jones et al., 2018). F1 heterozygote males were socially isolated for five days in an isolation chamber, paired with an unreceptive PMF for courtship LTM training and collected at various time-points during LTM formation, flash frozen using liquid nitrogen, heads isolated and stored at -80°C. Specifically, trained males were collected at three time-points during the courtship training period (DT; 1h, 3.5h, 7h) and at five time-points after training (AT; 1h, 4h, 7h, 13h, 19h). Naive male flies were also collected at five time-points (matching 1h-DT, 1h-AT, 7h-AT, 13h-AT, 19h-AT) to act as time-of-day controls. Each trained and naive cohort consisted of ~50 fly heads. All samples were processed for RNA-seq, except two 1h-DT naive samples, which were also processed to generate ATAC-seq libraries to look for constitutively accessible, regulatory regions of chromatin in the MB.

### 2.7.2 Bap60-KD collections

Male flies homozygous for either *GAL80<sup>ts</sup>;UAS-Bap60<sup>RNAi</sup>* or *GAL80<sup>ts</sup>;UAS-mCherry<sup>RNAi</sup>* were crossed to virgin females homozygous for *UAS-Unc84::GFP;R14H06-GAL4*. Crosses were established at 18°C. F1 heterozygote males (*Bap60*-KD or *mCherry*-KD) were collected after eclosion and raised for four days in an isolation chamber at 18°C and transferred to 29°C, to allow for RNAi knockdown and *UAS-Unc84::GFP* expression, one day prior to LTM training. Expression of the *UAS-Unc84::GFP* construct after one day at 29°C was confirmed through a standard upright microscope. Flies were paired with an unreceptive PMF for a seven-hour LTM courtship training period and collected one hour after training had ended. Naive flies for both *Bap60*-KD and *mCherry*-KD were collected at the same time as trained flies. Each trained and naive cohort, consisting of ~50 flies, were flash-frozen using liquid nitrogen, heads separated from the body by vortexing, followed quickly by separation using frozen sieves and stored at -80°C. All samples were processed to create RNA and ATAC-seq libraries from the same biological sample.

## 2.8 Isolation of nuclei tagged in a specific cell-type (INTACT)

To isolate mushroom body nuclei for downstream transcriptome and chromatin accessibility analyses, a modified, optimized version of the INTACT method was performed as previously described (Henry et al., 2012; Jones et al., 2018; Nixon et al., 2019). In general, the INTACT approach used here includes expression of UAS-Unc84::GFP, a *Caenorhabditis elegans*-derived tagged nuclear envelope protein, specifically in the MB using the GAL4 driver line *R14H06-GAL4* (Henry et al., 2012). This makes it possible to immunoprecipitate MB tissue-specific nuclei from a whole head homogenate by using anti-GFP antibody-bound beads. We have previously shown that transcript levels and chromatin accessibility in MB nuclei can be accurately profiled using this approach for INTACT (Jones et al., 2018).

To prepare for INTACT, antibody-bound beads were freshly prepared for each immunopurification by absorbing 5 µg of anti-GFP antibody (Invitrogen, G10362) to Protein G Dynabeads in PBS with 0.1% Tween-20 (PBST) for 10 minutes at room temperature. Simultaneously, beads without an antibody, used for non-specific pre-clearing, were washed in PBST for 10 minutes at room temperature. Antibody-bound and pre-clearing beads were then isolated using a magnet and resuspended in fresh PBST. Samples of approximately ~50 fly heads were suspended in 1mL of homogenization buffer ((25 mM KCl [pH 7.8], 5 mM MgCl<sub>2</sub>, 20 mM Tricine, 150 nM spermine, 500 nM spermidine, 10 mM β-glycerophosphate, 250 mM sucrose, 1x Halt protease inhibitor cocktail- EDTA-free [ThermoFisher Scientific: 78437], RNasin ribonuclease inhibitor (Promega: N2615) in a 1.5 mL Eppendorf tube, and head tissue crushed, and the suspension homogenized using a pestle. The suspension was then transferred to a Dounce homogenizer with NP-40 added to an end concentration of 0.3% and then homogenized six times with the tight pestle. The homogenate was filtered through a 40 µm strainer into a clean 50 mL Falcon tube and briefly spun down in a swinging bucket centrifuge cooled to 4°C. Here, specifically for all samples in the memory time-course (**Chapter 3**), an input whole-head (WH) nuclear fraction -containing both GFP-positive MB nuclei and non-GFP nuclei - was extracted by pipetting 50 µL of homogenate into a new Eppendorf tube, nuclei pelleted by centrifugation (4000xg, 10 minutes, 4°), supernatant discarded, and the pellet kept on ice. The remaining homogenate was then pre-cleared using non-antibody bound beads for 10 minutes at 4°C. Antibody-bound beads were added to the homogenate and incubated with rotation for 30 minutes at 4°C. Beads were then washed in homogenization buffer for 10 minutes at 4°C. Bead-bound

nuclei, containing MB-specific GFP nuclei, were then processed for either RNA-sequencing or ATAC-sequencing.

## 2.9 RNA-sequencing

Total RNA was isolated from the input WH nuclear fraction and the immunoprecipitated MB nuclei using the Arcturus PicoPure RNA isolation kit (ThermoFisher Scientific: KIT0204), and DNase digestion performed using the RNase-free DNase kit (Qiagen: 79254) according to the manufacturer's instructions. RNA quality was then assessed by visual examination of rRNA-peak integrity using the Bioanalyzer 2100 Pico RNA kit (Agilent: 5067-1513).

RNA-seq libraries were prepared using either the Nugen Ovation *Drosophila* RNA-Seq System 1-16 Kit or the Tecan Universal Plus Total RNA-seq library preparation kit according to manufacturer's instructions. Library size and quality was assessed with the Bioanalyzer 2100 DNA high-sensitivity kit (Agilent: 5067-4626). Sequencing was performed with the Illumina NovaSeq 6000 at Genome Quebec with the S4 v1.5 200 cycle kit; read length was 100 bp for paired-end reads.

## 2.10 ATAC-sequencing

ATAC-seq was performed as previously described (Buenrostro et al., 2015) with modifications for INTACT-isolated nuclei (Nixon, 2020). All ATAC-seq libraries were generated from one-third of INTACT-isolated bead-bound MB nuclei, containing ~50,000 nuclei, with the remaining two-thirds of MB nuclei processed into an RNA-seq library. This approach allowed for simultaneous profiling of transcript levels and chromatin accessibility in the MB from the same biological sample. Quality assessment insight into this approach is shown in **Appendix B** for data used in **Chapter 3**.

To generate ATAC-seq libraries, first, bead-bound MB nuclei were resuspended in ice-cold INTACT homogenization buffer and transferred into a new PCR tube. PCR tubes were placed on a magnet for five minutes and the supernatant discarded. Bead-bound nuclei were then resuspended in 50  $\mu$ L of transposase reaction mix (25 $\mu$ L 2XTD buffer, 2.5 $\mu$ L Tn5 Transposase (Illumina: 20034197), and 22.5  $\mu$ L of nuclease free water) and incubated for 30 minutes at 37°C in a thermal cycler. DNA was then purified (N.B. beads were placed directly onto the column) and eluted into 10  $\mu$ L of elution buffer using a Qiagen MinElute Kit according to the manufacturer's

instructions. Purified DNA was mixed with custom Nextera primers and High-Fidelity PCR Mastermix (NEB) and amplified for five cycles of 72°C for 5 minutes, 98°C for 30 seconds, (98°C for 10 seconds, 63°C for 30 seconds, 72°C for 1 minute) X 5, hold at 4°C. After five cycles libraries were removed from the thermal cycler and 5 µL was used in a qPCR side reaction to determine the number of additional cyclers for library amplification and help reduce GC, size bias and PCR duplication rate. Additional cycles were run for each library by determining the number of cycles corresponding to one-third of the maximum fluorescent intensity from the qPCR guide reaction. Amplified libraries were purified using a Qiagen PCR purification kit according to manufacturer's instructions and eluted into a volume of 20 µL elution buffer (10mM Tris Buffer, pH 8). Size selection was performed using Agencourt SPRIselect beads (Beckman-Coulter) to remove excess primers from the final libraries. Library size and quality was assessed using the Bioanalyzer DNA high-sensitivity kit (Agilent). Sequencing was performed with the Illumina NovaSeq 6000 at Genome Quebec with the S4 v1.5 200 cycle kit; read length was 100 bp for paired-end reads.

## **2.11 Bioinformatic analysis of sequencing data**

### **2.11.1 RNA-sequencing analysis**

RNA-seq reads were processed on Compute Canada servers (StdEnv/2020). First, raw reads were lightly trimmed, and adaptors clipped using Trimmomatic (v0.39) (Bolger et al., 2014). The read quality was then assessed using FastQC (v 0.11.9). Trimmed reads were aligned to the *Drosophila melanogaster* genome (Ensembl release 103, dm6) using STAR (v 2.7.5a) (Aken et al., 2016; Dobin et al., 2013). RNA-seq reads were also aligned to the *C. elegans Unc-84* gene (NC\_003284.9) to validate specificity of MB samples through comparison to WH samples. Uniquely aligned reads with a maximum of four mismatches were counted to genes using featureCounts or HTSeq-count (v 0.7.1) using the default union setting (Anders et al., 2015; Liao et al., 2014). The code for RNA-seq library processing can be found in **Appendix C**. Downstream analysis was done using RStudio (v 4.0.3), with differential expression analysis handled by DESeq2 (v 1.30.1) (Anders et al., 2010; Love et al., 2014; R Core Team, 2016). Further analysis of data, including gene annotation, gene ontology (GO), statistical comparison between groups of genes, was performed using elements of the R package BinTools (<https://github.com/kevincjnixon/BinTools>). Specific commands used included: count\_plot, getSym, barGene, zheat, GO\_GEM and customGMT. Data was further visualized using the R

packages ggplot2 (v 3.4.2) and pheatmap (v 1.0.12). Hypergeometric test statistics and overlap between datasets was determined in R, and Venn diagrams visualized using BioVenn(Hulsen et al., 2008; R Core Team, 2016).

### **Memory time-course analysis (Chapter 3)**

An average of 40,543,996 reads were generated across all MB (n=38) and WH (n=38) libraries generated (**Table 2.2-2.3**). Raw reads were filtered to select for those uniquely aligning to the *Drosophila melanogaster* genome with  $\leq 4$  mismatches. An average of 25,727,025 reads across all samples aligned to genes using featureCounts, for an average sequenced read to genic count efficiency of 63.45%. Counts were then filtered for rRNA and non-coding genes, genes mapped to the Y-chromosome or mitochondrial genome, and genes that had less than 150 normalized counts in 50 of MB and WH samples. After filtering, 6965 highly expressed genes were used for downstream differential expression analysis.

To identify memory regulated genes, differential expression was performed using DESeq2 and transcripts declared significant if FDR  $< 0.1$ . To identify genes specifically expressed in the WH or MB, differential expression was performed using DESeq2 and transcripts declared significantly MB or WH-enriched if FDR  $< 0.05$  and  $\log_2$  fold change  $> 0.5$  (MB) or  $\log_2$  fold change  $< -0.5$  (WH). Gene ontology analysis was performed using the BinfTools function GO\_GEM, with a background of all expressed *Drosophila* genes used, and terms declared significantly enriched if FDR  $< 0.05$ .

**Table 2.2 Read distribution for RNA-sequencing of INTACT-isolated MB samples**

<b>Sample Name</b>	<b>Total Reads</b>	<b>Unmapped</b>	<b>Good Reads<sup>a</sup></b>	<b>Genic Counts<sup>b</sup></b>
<b>1h-DT-N-1-MB</b>	45,797,102	6,830,397	33,369,220	29,760,061
<b>1h-DT-N-2-MB</b>	30,936,837	4,978,807	21,955,512	19,507,060
<b>1h-DT-N-3-MB</b>	43,315,711	6,547,863	31,339,384	27,732,824
<b>1h-DT-T-1-MB</b>	47,614,996	4,584,082	35,600,314	31,636,319
<b>1h-DT-T-2-MB</b>	41,721,053	6,074,655	31,440,128	27,979,691
<b>1h-DT-T-3-MB</b>	50,514,013	5,834,439	37,413,636	33,306,099
<b>3.5-DT-T-1-MB</b>	42,943,969	3,750,429	71,947,767	29,267,934
<b>3.5-DT-T-2-MB</b>	44,911,869	4,726,158	35,973,884	32,069,865
<b>3.5-DT-T-3-MB</b>	37,003,192	4,896,797	27,014,211	24,120,238
<b>7h-DT-T-1-MB</b>	50,874,894	3,992,077	39,531,269	35,249,772
<b>7h-DT-T-2-MB</b>	47,217,957	3,389,819	37,872,705	33,789,434
<b>7h-DT-T-3-MB</b>	43,559,858	4,538,454	33,276,754	29,773,415
<b>1h-AT-N-1-MB</b>	36,251,033	5,980,172	24,914,189	22,137,351
<b>1h-AT-N-2-MB</b>	34,750,327	4,453,074	25,616,558	22,726,584
<b>1h-AT-N-3-MB</b>	33,179,446	4,817,011	23,574,801	20,865,339
<b>1h-AT-T-1-MB</b>	35,296,179	3,125,671	26,836,815	23,904,830
<b>1h-AT-T-2-MB</b>	35,583,753	4,436,017	25,717,635	22,790,059
<b>1h-AT-T-3-MB</b>	38,038,242	4,310,614	27,489,506	24,346,868
<b>4h-AT-T-1-MB</b>	38,659,177	3,864,533	30,214,243	26,855,268
<b>4h-AT-T-2-MB</b>	41,347,258	3,083,239	32,848,633	29,422,218
<b>7h-AT-N-1-MB</b>	52,135,607	6,750,121	38,109,157	33,976,534
<b>7h-AT-N-2-MB</b>	41,368,347	6,079,699	29,103,245	25,955,777
<b>7h-AT-N-3-MB</b>	36,734,599	4,955,156	26,838,519	23,927,855
<b>7h-AT-T-1-MB</b>	108,619,108	11,348,630	83,615,579	74,825,469
<b>7h-AT-T-2-MB</b>	45,774,473	4,326,211	34,872,632	31,012,437
<b>7h-AT-T-3-MB</b>	48,388,864	4,984,256	37,433,152	33,178,219
<b>13h-AT-N-1-MB</b>	50,309,087	7,454,984	35,800,933	31,818,164
<b>13h-AT-N-2-MB</b>	36,688,833	3,348,567	27,430,059	24,449,975
<b>13h-AT-N-3-MB</b>	59,467,050	9,420,815	43,382,752	38,581,021
<b>13h-AT-T-1-MB</b>	37,560,136	4,835,710	27,865,990	24,770,092
<b>13h-AT-T-2-MB</b>	50,821,902	7,491,824	36,909,469	32,981,767
<b>13h-AT-T-3-MB</b>	37,321,158	2,419,832	29,577,532	26,430,414
<b>19h-AT-N-1-MB</b>	52,135,607	6,750,121	38,109,157	33,976,534
<b>19h-AT-N-2-MB</b>	41,368,347	6,079,699	29,103,245	25,955,777
<b>19h-AT-N-3-MB</b>	36,734,599	4,955,156	26,838,519	23,927,855
<b>19h-AT-T-1-MB</b>	108,619,108	11,348,630	83,615,579	74,825,469
<b>19h-AT-T-2-MB</b>	45,774,473	4,326,211	34,872,632	31,012,437
<b>19h-AT-T-3-MB</b>	48,388,864	4,984,256	37,433,152	33,178,219

<sup>a</sup>Good reads are uniquely aligned to the *Drosophila* genome with less than four mismatches

<sup>b</sup> Genic counts column is final read counts to genic regions using featureCounts used for downstream differential expression analysis

**Table 2.3 Read distribution for RNA-sequencing of INTACT-input WH samples**

<b>Sample Name</b>	<b>Total Reads</b>	<b>Unmapped</b>	<b>Good Reads<sup>a</sup></b>	<b>Genic Counts<sup>b</sup></b>
1h-DT-N-1-	35,800,441	10,952,112	20,602,997	17,988,795
1h-DT-N-2-	42,183,943	13,856,619	23,730,870	20,796,288
1h-DT-N-3-	35,822,083	7,259,189	24,007,217	21,003,444
1h-DT-T-1-	41,061,434	5,879,609	28,429,075	24,857,630
1h-DT-T-2-	46,899,315	13,570,303	27,164,490	23,731,273
1h-DT-T-3-	43,159,893	8,108,775	29,133,258	25,576,630
3.5-DT-T-1-	37,196,249	3,750,777	26,703,006	23,384,952
3.5-DT-T-2-	34,537,635	4,971,358	23,764,429	20,847,534
3.5-DT-T-3-	17,524,096	3,757,032	10,454,127	9,126,794
7h-DT-T-1-	51,492,473	4,428,955	39,472,139	34,771,263
7h-DT-T-2-	41,027,091	3,220,156	30,861,651	27,234,833
7h-DT-T-3-	31,750,022	4,371,500	22,920,538	20,243,573
1h-AT-N-1-	32,140,985	8,497,263	18,960,198	16,649,284
1h-AT-N-2-	37,238,492	10,961,324	22,614,608	19,922,415
1h-AT-N-3-	35,872,668	7,527,432	23,187,287	20,399,516
1h-AT-T-1-	37,154,608	3,138,625	28,262,412	24,960,704
1h-AT-T-2-	34,732,966	5,226,120	24,346,560	21,492,939
1h-AT-T-3-	34,053,306	4,310,741	25,229,657	22,210,258
4h-AT-T-1-	33,870,276	4,403,468	23,776,221	20,938,138
4h-AT-T-2-	25,112,605	1,855,922	19,116,211	16,867,152
7h-AT-N-1-	34,228,217	7,316,416	22,148,611	19,559,379
7h-AT-N-2-	31,097,814	8,505,164	18,969,842	16,672,764
7h-AT-N-3-	46,245,802	9,708,898	29,932,326	26,335,270
7h-AT-T-1-	41,327,415	4,504,488	31,410,027	27,789,976
7h-AT-T-2-	33,457,592	3,196,977	25,831,396	22,832,436
7h-AT-T-3-	49,388,390	6,472,938	37,418,950	33,104,947
13h-AT-N-1-	26,124,644	7,315,153	15,199,974	13,329,335
13h-AT-N-2-	38,135,085	5,499,110	25,695,736	22,547,178
13h-AT-N-3-	27,841,752	7,465,020	17,356,557	15,285,420
13h-AT-T-1-	42,097,827	7,087,623	30,591,642	27,008,882
13h-AT-T-2-	36,895,099	8,135,585	24,885,920	21,960,385
13h-AT-T-3-	36,961,932	2,288,165	28,805,170	25,327,869
19h-AT-N-1-	36,230,828	12,395,012	19,604,829	17,146,131
19h-AT-N-2-	42,197,996	8,250,690	28,489,293	24,978,844
19h-AT-N-3-	39,955,789	10,197,948	24,271,901	21,267,098
19h-AT-T-1-	36,972,508	4,764,645	26,722,646	23,371,625
19h-AT-T-2-	40,770,112	5,529,700	28,183,746	24,724,724
19h-AT-T-3-	34,710,186	4,842,959	23,923,319	20,953,336

<sup>a</sup>Good reads are uniquely aligned to the *Drosophila* genome with less than four mismatches

<sup>b</sup> Genic counts column is final read counts to genic regions using featureCounts used for downstream differential expression analysis

## Bap60-KD analysis (Chapter 4)

RNA libraries (n=8, duplicates for each condition) were sequenced to an average of 239,898,050 reads which were then filtered for those uniquely aligned to *Drosophila melanogaster* genome with  $\leq 4$  mismatches (Table 2.4). An average of 145,540,296 reads across all samples aligned to genes using HTSeq, for an average sequenced read to genic count efficiency of 60.6%. Counts were then filtered for rRNA and non-coding genes, genes mapped to the Y-chromosome or mitochondrial genome, and genes having the lowest one-third normalized expression levels. After filtering, 8626 highly expressed genes were used for differential expression analysis. Differentially expressed genes were defined as genes with an FDR of  $< 0.05$ . Gene ontology analysis was performed using the BinTools function GO\_GEM, compared to the background of highly expressed genes, terms declared significantly enriched if FDR  $< 0.05$ . RNA-seq tracks were visualized by normalizing bam files using the bamCoverage function from the command line package deepTools (v 3.5.1) with scale factors determined by DESeq2 (Love et al., 2014; Ram et al., 2016). Consensus track files were generated between replicates using the mean function from the command line program wiggletools (v 1.2) (Zerbino et al., 2014).

**Table 2.4 Read distribution for RNA-sequencing of Bap60-KD & mCherry-KD MB samples**

Sample Name	Total Reads	Unmapped/Multi-Mappers	Good Reads <sup>a</sup>	Genic Counts <sup>b</sup>
mCherry-KD-Naive-1	280,058,798	58,104,994	208,390,773	184,003,663
mCherry-KD-Naive-2	323,832,831	74,454,449	234,856,346	207,201,627
mCherry-KD-LTM-1	287,919,676	81,950,075	193,509,328	169,962,981
mCherry-KD-LTM-2	330,003,515	79,020,816	235,909,952	206,647,002
Bap60-KD-Naive-1	136,403,880	59,643,873	72,483,788	61,432,302
Bap60-KD-Naive-2	261,648,283	70,948,903	178,794,575	156,701,293
Bap60-KD-LTM-1	116,906,971	38,725,837	73,702,700	63,897,648
Bap60-KD-LTM-2	182,410,448	43,987,277	130,081,904	114,475,853

<sup>a</sup>Good reads are uniquely aligned to the *Drosophila* genome with less than four mismatches

<sup>b</sup> Genic counts column is final read counts to genic regions using featureCounts used for downstream differential expression analysis

### 2.11.2 ATAC-seq analysis

Sequenced ATAC-seq reads were processed on Compute Canada servers (StdEnv/2020). Reads were first lightly trimmed, and adaptors clipped using Trimmomatic (v 0.39). Trimmed reads were aligned to the *Drosophila melanogaster* reference genome (Ensembl release 103, dm6) using bowtie2 (v 2.4.1) with the settings -X 2000 and -very-sensitive (Aken et al., 2016; Bolger et al., 2014; Langmead & Salzberg, 2012). Reads aligning to multiple loci, the mitochondrial genome, and scaffolds were filtered using samtools view (v 1.11) (H. Li et al., 2009). Duplicate reads resulting from PCR amplification were identified using samtools fixmate and removed using samtools markdup. Reads were shifted, +4 bp for the forward strand and -5 bp for the negative strand, to account for the 9-bp duplication created by the DNA repair nick of the Tn5 transposase (Yan et al., 2020). The code for ATAC-seq library processing can be found in **Appendix D**.

Distribution of ATAC-seq fragment sizes was determined using bamPEFragmentSize from deepTools. Peaks were called using MACS2 software (v 2.1.2) using the settings -q 0.01 -min-length 50 and -max-gap 100 (Zhang et al., 2008). Peaks refer to regions of the genome where more sequencing reads align compared to the surrounding genomic background, and in the context of ATAC-seq represents an area of significantly accessible chromatin. Peaks were analyzed using RStudio (v4.0.3). Differential peak analysis, to find regions of significantly different chromatin accessibility between conditions, was performed using DiffBind (v 3.0.15), with libraries normalized using the total number of reads and the fraction of reads in peaks calculated (FRiP) (Stark & Brown, 2011). Samples had to have a FRiP of >0.3 for inclusion, as per ENCODE standards (Luo et al., 2020). Differential peaks were annotated using the R Package ChIPseeker (v 1.26.2) (Wang et al., 2022; Yu et al., 2015).

Promoter and genomic regions, for the purpose of creating BED files for downstream visualization, were extracted using the R annotation package TxDb.Dmelanogaster.UCSC.dm6.ensGene (v 3.12.0) in combination with GenomicRanges (v 1.42.0). For visualization, bam files were normalized using the bamCoverage function from deepTools with scale factors determined by the dba.normalize function from Diffbind. Consensus track files were generated between replicates using the mean function from the command line program wiggletools. Bandplot files for BED region subsets were generated using the computeMatrix and plotProfile functions from deepTools (Ram et al., 2016).

### **Constitutively accessible chromatin (Chapter 3)**

ATAC-seq libraries were generated in duplicate from ~50 naive male flies at a time-point corresponding to 1h after memory training onset. Libraries were sequenced to a depth of 85,293,731 and 109,182,398 reads, respectively. After removing reads aligning to multiple loci, the mitochondrial genome, and scaffolds, 38,388,868 and 42,678,219 high-quality reads remained for downstream chromatin accessibility visualization, respectively. Both libraries had a FRiP >0.3, as calculated by DiffBind. Peak calling resulted in 15842 stringent peaks identified uniquely between both samples, with 11705 consensus peaks – which represent regions of constitutively accessible chromatin or regulatory regions. Consensus peaks annotated to 7488 genes.

### **Bap60-KD (Chapter 4)**

ATAC-seq libraries (n=8, duplicates for each condition) were sequenced to an average depth of 144,356,581 reads. After removing reads aligning to multiple loci, the mitochondrial genome, and scaffolds, an average of 45,827,084 high quality reads remained for downstream chromatin accessibility analysis (**Table 2.5**). We called peaks ( $p < 0.01$ ) in all samples to identify specific regions in the MB with significant accessibility. We identified 20828 unique peaks between all samples, with 16293 consensus peaks, which we refer to as regulatory regions, found in at least two samples that were used for downstream differential accessibility analysis. Samples had a FRiP ratio ranging from 0.38 – 0.55, meeting the ENCODE standard minimum of 0.3. Differential peak analysis was performed using DiffBind, with libraries normalized using the total number of reads. Peak regions were determined to be differentially accessible if they had an FDR <0.05 between comparisons.

**Table 2.5 Read distribution for ATAC-sequencing of Bap60-KD & mCherry-KD MB samples**

<b>Sample Name</b>	<b>Total Reads</b>	<b>Unmapped</b>	<b>Good Reads<sup>a</sup></b>	<b>Peaks Called<sup>b</sup></b>	<b>FriP<sup>c</sup></b>
<b>mCherry-Naive-1</b>	339,287,155	41,480,658	133,163,624	17,942	0.53
<b>mCherry-Naive-2</b>	133,348,550	20,141,602	57,951,806	15,248	0.54
<b>mCherry-LTM-1</b>	67,019,324	13,835,740	17,851,144	13,822	0.55
<b>mCherry-LTM-2</b>	128,572,755	23,105,399	8,289,878	16,446	0.54
<b>Bap60-KD-Naive-1</b>	80,205,235	13,239,157	33,454,388	14,229	0.42
<b>Bap60-KD-Naive-2</b>	161,812,232	17,459,769	72,264,004	16,142	0.54
<b>Bap60-KD-LTM-1</b>	152,755,120	27,627,931	19,120,524	12,344	0.48
<b>Bap60-KD-LTM-2</b>	91,852,275	25,313,946	24,521,303	17,827	0.38

<sup>a</sup>Good reads are uniquely aligned to the *Drosophila* genome with less than four mismatches after the removal of mitochondrial and duplicate reads.

<sup>b</sup>Peaks called column is the number of significant peaks called by MACS2 (FDR <0.01).

<sup>c</sup>FriP is the measure of total reads that are found within the significantly called peaks and is a measure of ATAC-seq quality, with a value of 0.3 deemed high-quality by ENCODE standards.

## 2.12 ChIP-seq analysis

To identify binding sites for CrebB, as well as the transcription factors Hr38 and sr, publicly available ChIP-seq data was obtained from the ENCODE project repository (ENCODE Project Consortium, 2012; Luo et al., 2020). Specifics regarding biological condition and source, as well as exact files used for binding site analysis and track visualization are available in **Table 2.6**. To determine genes with significant binding sites for either Hr38, sr or CrebB, bed files containing regions of optimal IDR thresholded peaks generated using the ENCODE pipeline, representing at least two or more biological replicates, were annotated to the nearest gene using `annotatePeaks.pl` from HOMER (v 4.11) (Heinz et al., 2010). CrebB or sr binding signal was visualized using `pyGenomeTracks` (v 3.6), using either control normalized or signal p-value bigWigs (Lopez-Delisle et al., 2021).

**Table 2.6 Transcription factor ChIP-seq data obtained from ENCODE**

Gene	Accession	File type	Description
<b>CrebB</b>	ENCFF090JJN	Bed narrowPeak, optimal IDR thresholded peaks from three biological replicates	ChIP-seq using whole organism embryo (0-24 hours) expressing CrebB-eGFP fusion proteins with an anti-GFP antibody.
<b>CrebB</b>	ENCFF655EMQ	bigWig, control normalized signal from combined three biological replicates	
<b>Hr38</b>	ENCFF144OZH	Bed narrowPeak, optimal IDR thresholded peaks from two biological replicates	ChIP-seq on 48-hour whole organism prepupa expressing a Hr38-eGFP fusion protein with an anti-GFP antibody.
<b>sr</b>	ENCFF186BCY	bigWig, signal p-value from three combined biological replicates	ChIP-seq on 8–24-hour whole organism embryo expressing a sr-eGFP fusion protein with an anti-GFP antibody.
<b>sr</b>	ENCFF247KLE	Bed narrowPeak, optimal IDR thresholded peaks from three biological replicates	

### 2.13 MB-specific Brm-TaDa

Targeted DamID (TaDa) is a genomics approach that enables the cell-type specific profiling of DNA-binding proteins with fine temporal and spatial resolution (Marshall et al., 2016; Southall et al., 2013). TaDa is a technique built upon DamID (DNA adenine methyltransferase identification). In this approach, a bacterial DNA adenine methylase (Dam) is fused to a protein of interest (Steensel & Henikoff, 2000). Consequently, only the genomic regions bound by the protein will undergo methylation at the GATC sequence. The methylated GATC fragments are then selectively amplified using methylation-sensitive digestion and ligation-mediated PCR. Subsequent sequencing of these fragments allows for the determination of the binding profile of the protein of interest. TaDa is a method that merges DamID with the GAL4 system to enable the expression of Dam-fusion proteins exclusively in specific cell types during precise temporal windows.

To determine SWI/SNF binding sites in the MB during LTM formation we obtained TaDa data generated by our collaborator, Dr. Francisco Martin-Castro (Autonomous University of Madrid). Here, a fly was created with the core ATPase of the *Drosophila* SWI/SNF complex, *brm*, fused to Dam to generate the genotype *UAS-Dam-brm* (Brm-TaDa). Brm-TaDa was expressed specifically in the MB using the GAL4 driver *MB247-GAL4* under the temperature-sensitive control of GAL80<sup>ts</sup>. Male flies were maintained at 17°C for five days, which limited Brm-TaDa methylation activity, courtship memory trained by pairing with an unreceptive PMF for five hours at 25 °C, and then immediately moved to 29°C post-training to activate Brm-TaDa and begin methylating SWI/SNF bound DNA. Flies were left at 29°C to profile *brm*-binding for 20h, then flash frozen and processed for sequencing. Processed Brm-TaDa bedgraph files were obtained for trained flies for downstream analysis. Peaks were called using the freely available command line script `find_peaks` ([https://github.com/owenjm/find\\_peaks](https://github.com/owenjm/find_peaks)), with an FDR significance cut-off of 0.01. Peaks were then annotated to the nearest gene, using the script `peaks2genes`, using the default settings. Genes annotated with significant Brm-TaDa binding we classified as likely SWI/SNF-bound memory genes. Brm-TaDa binding BED files, along with RNA, ATAC and sr ChIP-seq profiles were plotted together using pyGenomeTracks for a subset of candidate Bap60-dependent memory genes (Lopez-Delisle et al., 2021).

## Chapter 3: The transcription factor hierarchy activated during long-term memory formation

### 3.1 Validating established memory formation dogma in the long-term courtship suppression memory assay

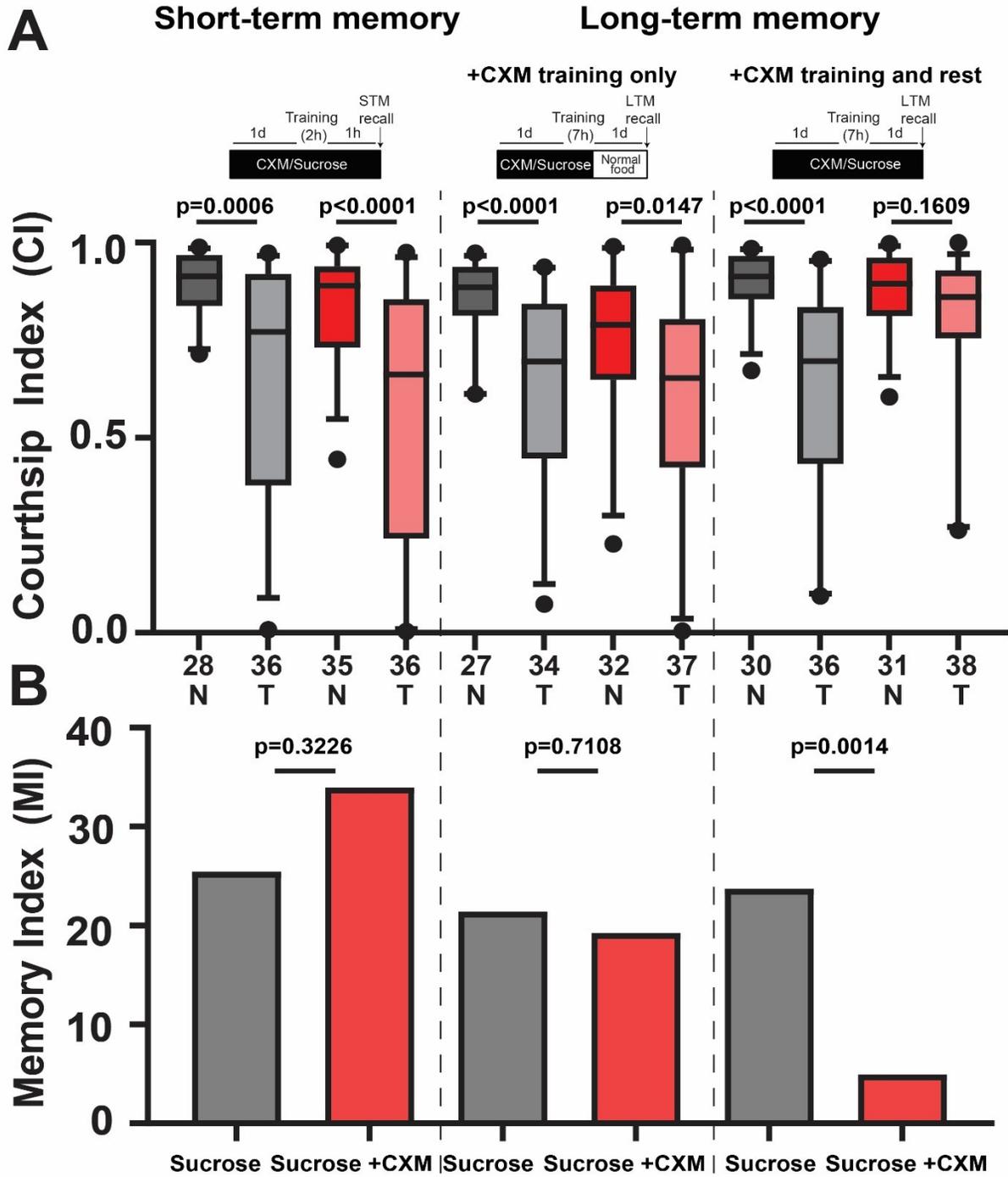
The requirement for translation and CREB-mediated transcription during LTM formation has been firmly established in *Drosophila* using conditioned odor-avoidance and appetitive memory assays, however, recent evidence has revealed that a context-dependent LTM formed by aversive olfactory conditioning exists that is protein synthesis independent (Hirano et al., 2016; Yin et al., 1994; Zhao et al., 2019). This suggests that core principles of LTM formation may vary between different forms of memory. Previously, we showed that long-term courtship memory formation was associated with an increase in gene transcription in the MB after training (Jones et al., 2018), however, to date there has been no previous work that shows a general requirement for translation during courtship memory formation. This study looks to elucidate the temporal dynamics of transcription during long-term courtship memory formation. However, as the main argument for studying the role of transcription during LTM is that downstream protein translation of training induced transcripts is necessary for formation it is important to confirm that long-term courtship memory formation requires translation.

To test the requirement for translation during long-term courtship memory formation protein synthesis was blocked by feeding flies the chemical cycloheximide. Cycloheximide blocks protein synthesis by interfering with translocation in the ribosome and has been used previously in *Drosophila* to show that translation is required in long-term conditioned odor-avoidance memory (Tully et al., 1994). In the following experiments (**3.1.1 and 3.1.2**), distinct temporal requirements for translation during long-term courtship memory formation are shown and it is demonstrated that CrebB is a critical transcription factor required in the MB for courtship memory formation.

### 3.1.1 Translation is required for long-term courtship memory but is dispensable during training and short-term memory formation

To identify a requirement for translation during courtship memory formation, protein synthesis was inhibited by feeding wild-type male flies media containing sucrose and cycloheximide (sucrose + CXM) beginning one day prior to STM and LTM training. Sucrose + CXM feeding continued throughout training, during the rest period, and up until memory-recall testing. Control flies were fed a sucrose-only diet in parallel to flies being fed sucrose + CXM. For STM, it was observed that both sucrose and sucrose + CXM trained males courted significantly less than naive males ( $p < 0.0001$ ,  $p = 0.0006$ , respectively) (**Figure 3.1A, left panel**), and showed no significant difference in MI ( $p = 0.3226$ ) (**Figure 3.B, left panel**). For LTM, the CI of trained flies was significantly reduced compared to untrained naive males when fed the control sucrose diet ( $p < 0.0001$ ), however, there was no significant reduction in CI of trained males fed sucrose + CXM when compared to naive ( $p = 0.1609$ ) (**Figure 3.1A, right panel**). This is reflected in a significant reduction of LTM MI between sucrose and sucrose + CXM flies (**Figure 3.1B, right panel**). This suggests that translation is indeed required during long-term courtship memory formation.

To dissect the temporal requirements for translation during LTM formation, timing of sucrose + CXM feeding was modified to allow for translation to resume following LTM training. Here, sucrose + CXM feeding commenced one day prior to training, with flies trained on CXM, but then returned to standard media during the rest period. Interestingly, trained males were found to court significantly less than naive males on both the sucrose and sucrose + CXM diet ( $p < 0.0001$ ,  $p = 0.0147$ , respectively) (**Figure 3.1A, middle panel**), and showed no significant difference in MI ( $p = 0.7108$ ) (**Figure 3.1B, middle panel**). Taken together, these results suggest that translation is required for LTM formation but is dispensable during training.



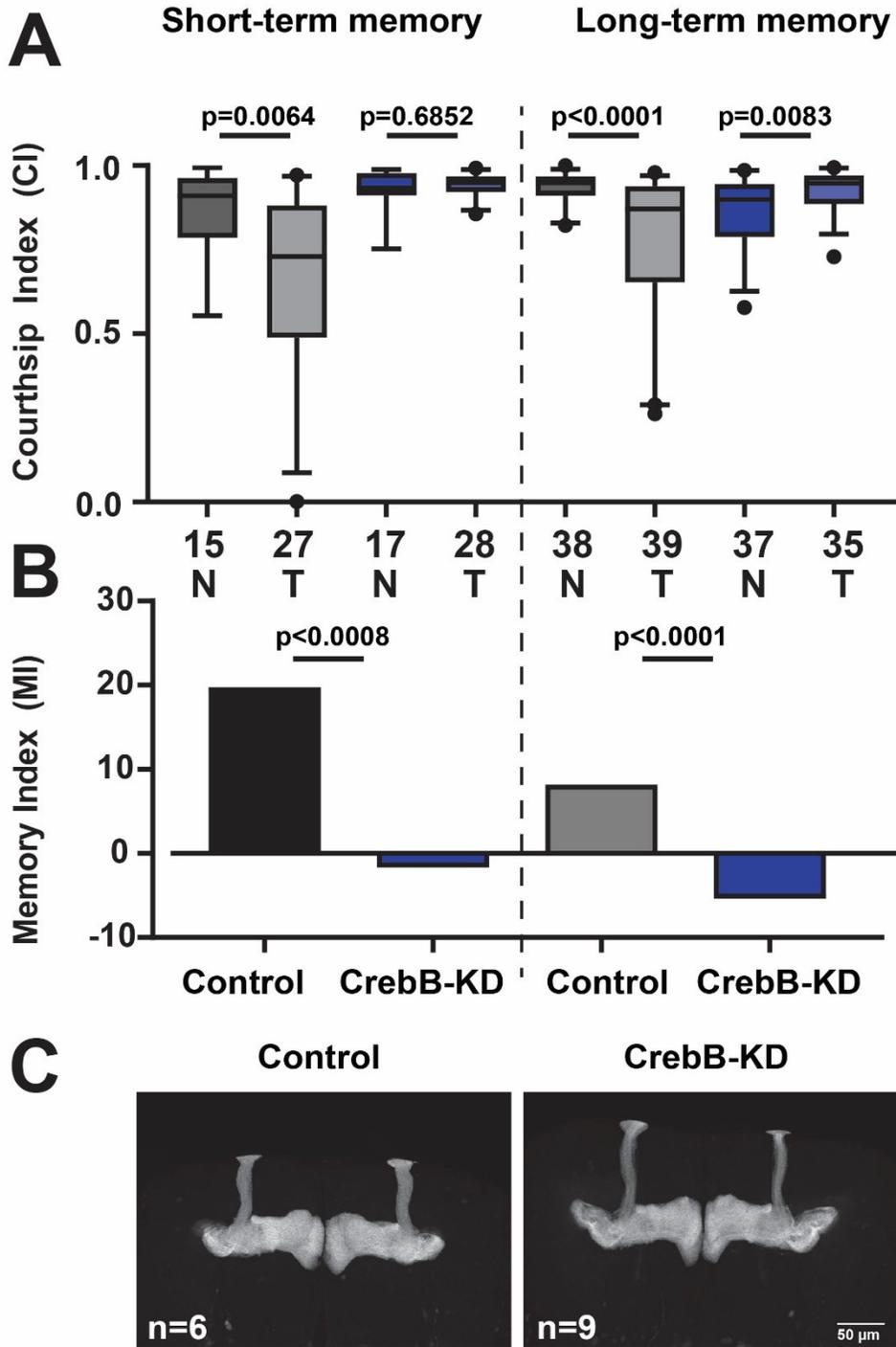
**Figure 3.2 Cycloheximide-mediated translation inhibition prevents LTM, but not STM formation**

(A) Boxplots distribution of courtship indices (CI) for naive (N) and trained (T) flies fed either a sucrose (grey) or sucrose + cycloheximide (CXM, red) diet. Statistical comparison was made between naive and trained flies for each feeding and training protocol, showing that only trained flies fed the sucrose + cycloheximide (CXM) diet throughout LTM training and up until memory recall were not capable of forming LTM (Mann-Whitney test). Whiskers represent 5-95% percentiles. Dashed line separates distinct cycloheximide feeding and training protocols. Total number of flies represented on the row below the x-axis. Schematic showing the feeding of CXM indicated above the boxplots. (B) Comparison between control and translation-inhibited condition memory indices (MI) show a significant reduction in MI specifically for sucrose + CXM fed during and after LTM training, with exact p-values shown above each comparison (randomization test, 10,000 bootstrap replicates).

### 3.1.2 *CrebB* is required in the MB for STM and LTM formation

*CrebB* is a member of the CREB transcription factor family and is a constitutively expressed transcription factor in the MB canonically required for LTM-induced transcription, a role which is conserved across several *Drosophila* memory assays including associative courtship memory (Sakai et al., 2004; Yin et al., 1994). Here, *CrebB* was tested for a role in cVA-retrievable courtship memory following RNAi-knockdown in the MB (*CrebB*-KD). CI was found to be significantly reduced in trained control flies ( $p=0.0064$ ) compared to naive males but was not reduced in *CrebB*-KD ( $p=0.6852$ ) (**Figure 3.2A, left panel**). Accordingly, the STM MI was significantly reduced in *CrebB*-KD flies when compared to controls ( $P<0.0008$ ) (**Figure 3.2B, left panel**). For LTM, *CrebB*-KD trained flies showed a significant increase in courtship activity in comparison to naive flies ( $p=0.0083$ ), with trained control flies showing a significant reduction in courting ( $p<0.0001$ ) (Figure x A, right panel). As expected, *CrebB*-KD LTM MI performance was significantly reduced in comparison to controls ( $p<0.0001$ ) (**Figure 3.2B, right panel**). Taken together, this evidence suggests that *CrebB* is required for normal formation of both STM and LTM in the MB.

Normal MB development is required for proper learning and memory processes to occur. With an unexpected STM phenotype, a process which is transcription-independent, *CrebB*-KD was investigated for a role in establishing normal MB morphology. Using confocal microscopy, no gross morphological defects were observed in the MB of *CrebB*-KD flies ( $n=9$ ) when compared to the control genetic background ( $n=6$ ) (**Figure 3.2C**). This suggests that *CrebB*-KD MB memory deficiency phenotypes are not due to developmental structural deformities in the MB and instead likely reflects a critical role for *CrebB* as a transcription factor during memory processes.



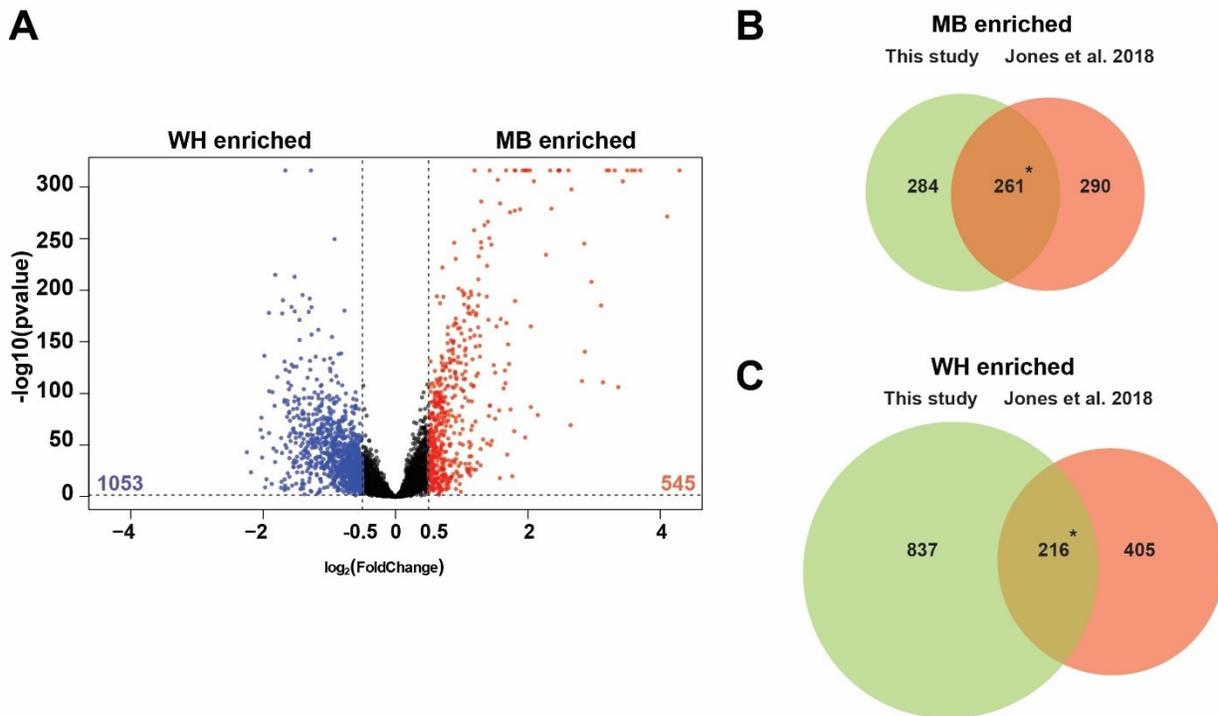
### Figure 3.3 MB-specific knockdown of *CrebB* causes defects in STM and LTM

**(A)** Boxplot distribution of courtship indices (CI) for naive (N) and trained (T) flies of RNAi-mediated knockdown of *CrebB* in the MB (*CrebB*-KD, blue) and genetic background controls (red) for short-term (STM) and long-term memory (LTM) assays. Statistical comparisons were made between naive and trained flies revealing *CrebB*-KD inhibits the formation of both STM and LTM (Mann-Whitney test). Whiskers represent 5-95% percentiles. Total number of flies represented on the row below the x-axis. **(B)** Comparison between control and *CrebB*-KD memory indices (MI) shows a significant reduction in MI for both STM and LTM, with exact p-values shown above each assay (randomization test, 10,000 bootstrap replicates). **(C)** Confocal images of *CrebB*-KD in the MB (MB highlighted using UAS-mCD8::GFP driven by R14H06-GAL4) show no visible structural defects. Scale bars: 50  $\mu\text{m}$ . Dissections and confocal microscopy were conducted by Olivia Kerr.

### 3.2 INTACT obtains MB-specific nuclear RNA for profiling a time-course of LTM transcription

Having established that translation is required for LTM, but not STM, as well as identifying a critical role for the transcription factor CrebB in the MB to establish cVA-dependent courtship memory, we next looked to better understand the transcriptional memory trace induced in MB memory-specific tissue that could be translated and required for establishing normal LTM. Here, we aimed to provide novel insight into the molecular mechanisms underlying training induced transcription by asking two main questions: 1) what are the genes induced by memory training in the MB and 2) how are these memory sensitive genes temporally regulated while LTM is forming?

To identify transcriptional changes during LTM formation in the MB, our previously optimized INTACT protocol was used to profile nuclear transcript level changes at eight time-points after courtship training onset in memory-associated MB and whole head (WH) nuclear extracts (Jones et al., 2018). To first assess the tissue specificity of INTACT to isolate MB-enriched nuclei, differential expression was performed between all WH (n=38) and MB (n=38) samples. Transcripts were identified that were significantly and highly enriched in each nuclear subset (FDR <0.05, log<sub>2</sub> fold change +/- 0.5). This approach revealed 1053 WH and 545 MB-enriched transcripts (**Figure 3.3A**) (**Appendix E**). GO analysis of WH-enriched genes included significant enrichment of sensory terms like “odorant binding” (ex. *Obp19d*, *Obp44a*, *Obp99c*) and “detection of visible light” (ex. *ninaE*, *ninaC*, *trp*, *rdgA*, *inaD*), which are associated with sensory organs like the eye and not with memory-relevant tissue. Significant MB-enriched gene terms were memory-relevant and included “postsynaptic neurotransmitter receptor activity” (ex. *nAChRalpha6*, *nAChRalpha5*, *Dop2R*, *Dop1R1*), cAMP-mediated signaling (ex. *Pka-C1*, *dnc*, *Pka-R2*, *Pka-R1*) and “anesthesia-resistant memory” (ex. *rut*, *rad*, *Rgk1,5-HT1A*). Further, MB and WH-enriched transcripts identified here were highly correlated with the WH and MB-enriched transcripts we identified previously (Jones et al., 2018), with a significant overlap of 261 and 216 genes, respectively (hypergeometric p-value < 0.05) (**Figure 3.3B-C**). Collectively these results provide evidence that INTACT was effectively used in this study to specifically isolate MB nuclei.



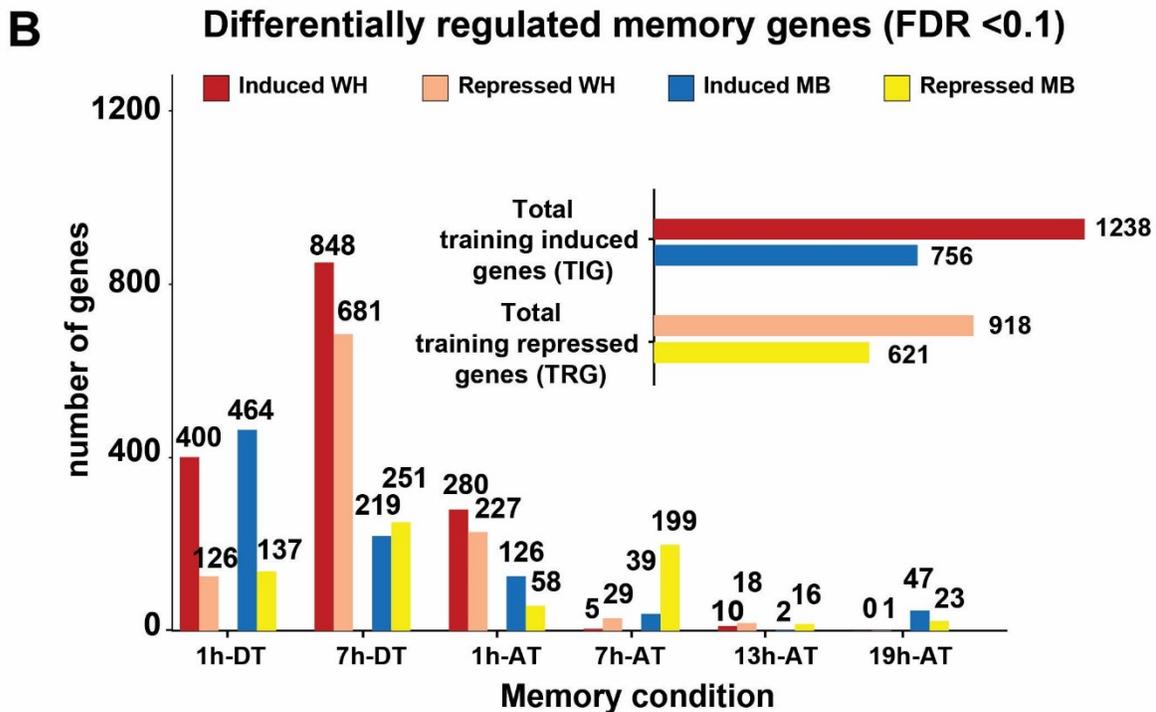
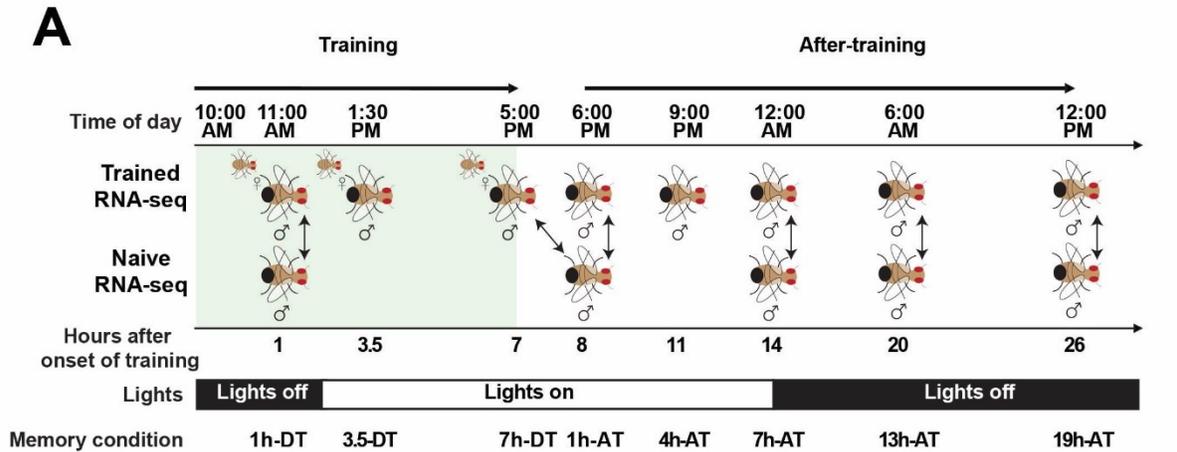
**Figure 3.4 Visualization of MB and WH-enriched differentially expressed genes and overlap with previously published tissue-specific genes**

**(A)** Volcano plot of differential expression analysis results between all whole-head (WH) (n=38) and INTACT-obtained mushroom-body (MB) samples (n=38). 6965 genes were used for differential expression analysis. Differentially expressed genes (FDR<0.05) MB-enriched are highlighted in red ( $\log_2$  fold change >0.5) and WH-enriched genes are highlighted in blue ( $\log_2$  fold change <-0.5). **(B)** Overlap between MB-enriched genes identified in this study and previously identified MB-enriched genes (Jones et al., 2018) reveals a significant degree of overlap (\*-hypergeometric test  $p=1.17 \times 10^{-160}$ , fold change = 6.01). **(C)** Overlap between WH-enriched genes found in this study and WH-enriched genes identified previously (Jones et al., 2018) reveals a significant degree of overlap (\*-hypergeometric test  $p=1.06 \times 10^{-37}$ , fold change = 2.29).

### 3.3 Identifying transcriptionally regulated genes during LTM formation

To identify genes regulated during LTM formation differential expression analysis was performed for both MB and WH fractions by contrasting trained flies with time-of-day matched controls at six different time-points: 1h and 7h during training (DT), and 1h, 7h, 13h and 19h after training (AT) (**Figure 3.4A**). Transcripts were declared significant if FDR <0.1 and classified as training induced genes (TIG) or training repressed genes (TRG) if transcript levels increased or decreased, respectively, at any time-point compared.

Overall, more unique genes were found that were differentially regulated in the WH fraction (2066) compared to the MB (1256). In both nuclear subsets more genes were induced than repressed, with 1238 TIG and 918 TRG identified in the WH, and 756 TIG and 621 MB TRGs (**Figure 3.4B**). The most transcriptional activity was observed during the courtship training period, which has not been previously profiled, with 1891 WH and 985 MB memory training-sensitive genes being identified at these two time-points, representing nearly 90% of all TIGs and TRGs. Taken together, this suggests a large network of genes are transcriptionally regulated during the courtship training period that could be relevant for LTM formation.



**Figure 3.5** Differential expression analysis between trained and naive flies during a time-course of LTM formation

(A) Schematic of courtship training and sample collection approach for both trained and naive flies. Arrows represent trained-naive comparisons used in downstream differential expression analysis. (B) Differential expression analysis result, separated by upregulation or downregulation and nuclear subset, for each trained-naive comparison (FDR <0.1).

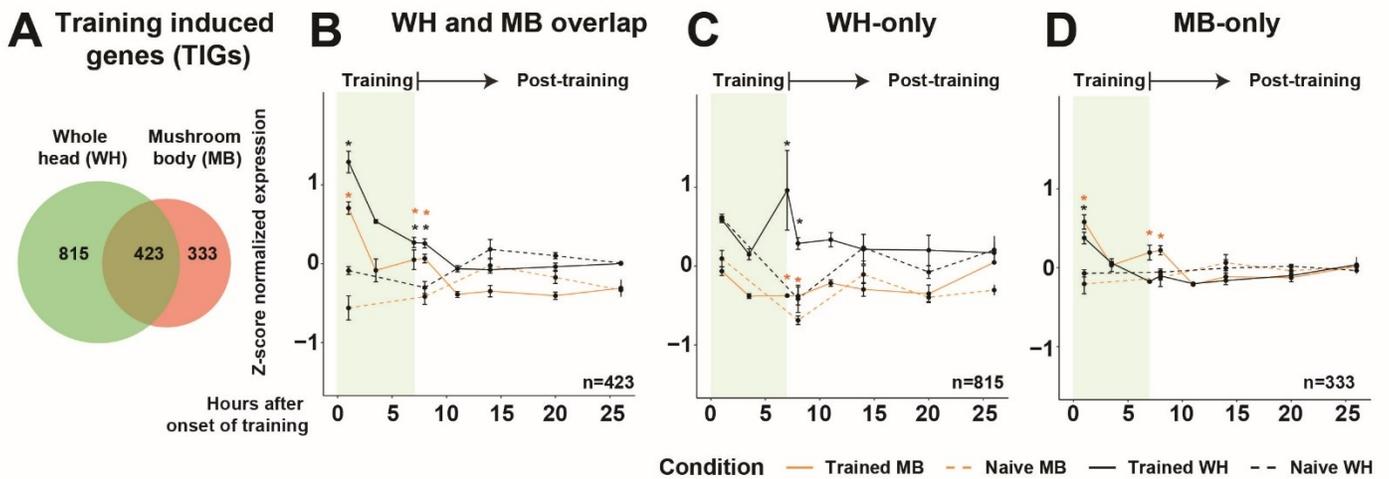
### 3.4 MB-specific transcript induction occurs as memory training progresses

We next looked to identify temporal trends in the transcription of WH and MB TIGs during LTM formation. First, WH and MB TIGs were overlapped to help distinguish the transcriptional trace specific to MB memory-relevant tissue, WH memory non-specific tissue, or common to both tissues. 423 TIGs were identified in both the WH and MB, 815 identified specifically in the WH, and 333 TIGs found only in the MB (**Figure 3.5A**) (**Appendix F**). WH and MB normalized expression was then plotted for each group of TIGs across the transcriptional time-course and significant induction determined.

In the overlapping WH and MB TIGs (n=423), there was significant induction 1h-DT, with transcript levels subsiding as training persists but maintaining significantly elevated transcription in both WH and MB tissue at the 7h-DT and 1h-AT time-points (**Figure 3.5B**). While the transcriptional trend of these overlapping TIGs is similar in both the WH and MB, there is slightly higher expression in the WH.

In WH-only TIGs (n=815) no transcription above baseline 1h-DT was observed (**Figure 3.5C**). Instead, significant expression of WH-specific TIGs occurs 7h-DT and 1h-AT in both the WH and MB. However, this later induction of WH-only TIGs is strongly above the naive baseline specifically in WH samples, with only a slight increase in the MB. Interestingly, transcription of WH-only TIGs in control naive flies is observed to follow a cyclical pattern which peaks when the incubator lights turn on or off (2h-DT, 7h-AT & 19h-AT).

Finally, in the 333 MB-only TIGs, like in overlapping TIGs identified in both the WH and MB, significant transcript induction was observed 1h-DT in both the WH and MB. Transcript levels then subside to near baseline levels 3.5h-DT (**Figure 3.5D**). Significant re-induction of MB-only TIGs was observed 7h-DT and 1h-AT only in the MB and not in the WH. These genes do not display any cyclical transcription in controls. Taken together, these results suggest there is a transcriptional program activated at the end of courtship training and early after separation specifically in memory tissue and that it could play a critical role in LTM formation.



**Figure 3.6 Training induced gene transcription becomes more divergent as courtship training progresses**

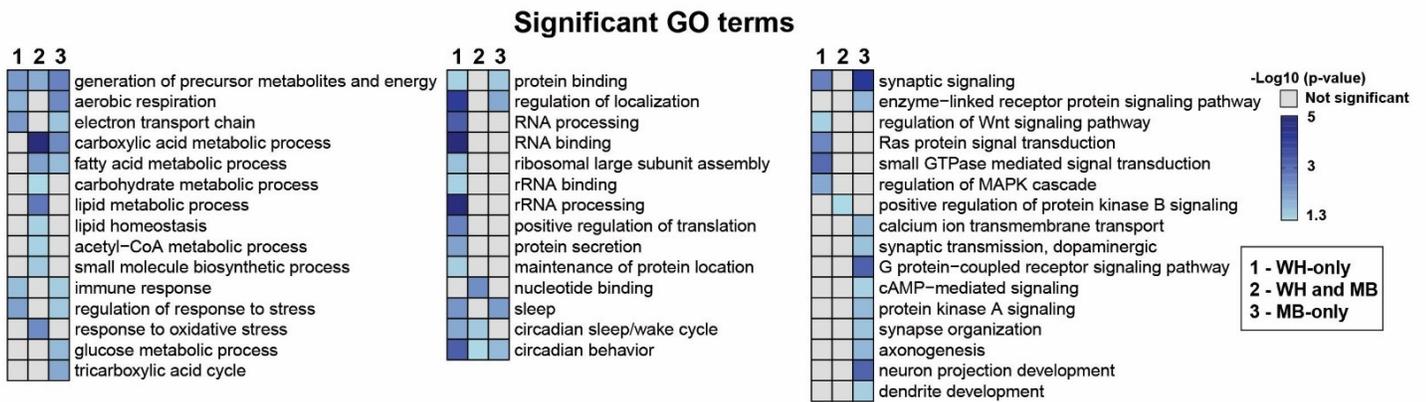
(A) Venn diagram of training induced genes (TIGs) identified in whole-head (WH) and mushroom body (MB) tissue. Mean z-score normalized expression during LTM formation for (B) TIGs identified in both the WH and MB (C) WH-only or (D) MB-only. Expression was compared between trained, and time-of-day matched naive flies to identify significant induction at time-points noted on x-axis (pairwise t-test, \* =  $p < 0.05$ ).

### 3.5 Metabolism and memory signaling are key components of the MB-specific memory transcriptional response

GO analysis was performed on TIGs identified in both the WH and the MB, or only in the WH, or only in the MB. The most prominent functionally enriched terms among TIGs identified in both the WH and the MB, which had 186 significant terms overall, were related to metabolism. Specifically, genes annotated to biological processes including “carbohydrate metabolic process” and “lipid metabolic process”, which were not found to be enriched in other TIG groups (**Figure 3.6, left column**).

TIGs identified only in WH had 612 significantly enriched terms, which were characterized by functions related to translation, circadian regulation, and sleep. Here, we identified genes annotated to terms including “positive regulation of translation, “RNA processing” and “rRNA binding”, which were only significant in the WH-only TIG group (**Figure 3.6, middle column**). WH-only TIGs also were annotated to terms like “sleep”, “circadian behaviour” and “immune response”. These terms were interestingly also found to be significantly enriched in MB-only TIGs.

In the TIGs that were identified only in the MB, significant enrichment was found for 464 terms and included functions related to metabolism and signalling pathways. Specifically, terms associated with “tricarboxylic acid cycle” and “glucose metabolic process” were enriched, which were only significant among MB-only TIGs (**Figure 3.6, left column**). Terms associated with known memory-relevant signalling pathways were also found to be enriched specifically among MB-only TIGs, including “cAMP-mediated signaling” and “synaptic transmission, dopaminergic” (**Figure 3.6, right column**). Taken together, functional enrichment of TIGs identified only in the MB reveal genes with established roles in LTM, suggesting that other genes part of the transcriptional program identified as training progresses in memory tissue may also play fundamental roles in LTM.



**Figure 3.7 Gene ontology analysis of training induced genes in different tissue-subsets during LTM formation**

Gene ontology analysis was performed for groups of training induced genes (TIGs) identified during a time-course of LTM formation. Significant GO terms (FDR <0.05) were identified for biological processes, molecular functions, and cellular components by contrasting groups of TIGs identified in both the WH and MB, WH-only or MB-only to a background of all *Drosophila* genes. A selection of curated terms is presented with  $-\text{Log}_{10}(\text{p-value})$  shown on a scale, with darker blue squares representing greater significance of that term in the corresponding TIG tissue subset.

### 3.6 *Hr38* and *sr* are candidate courtship memory IEGs induced during long-term courtship memory

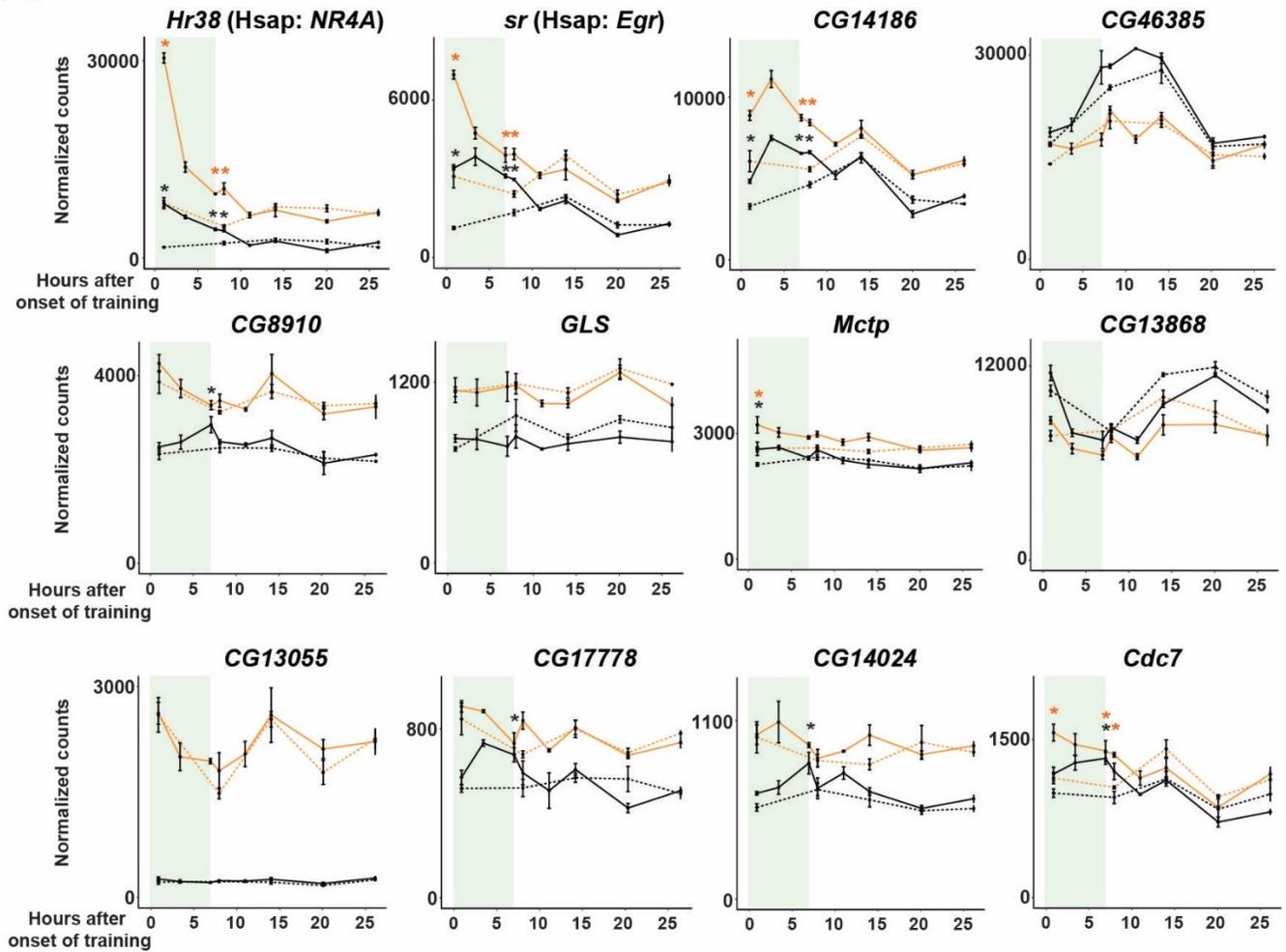
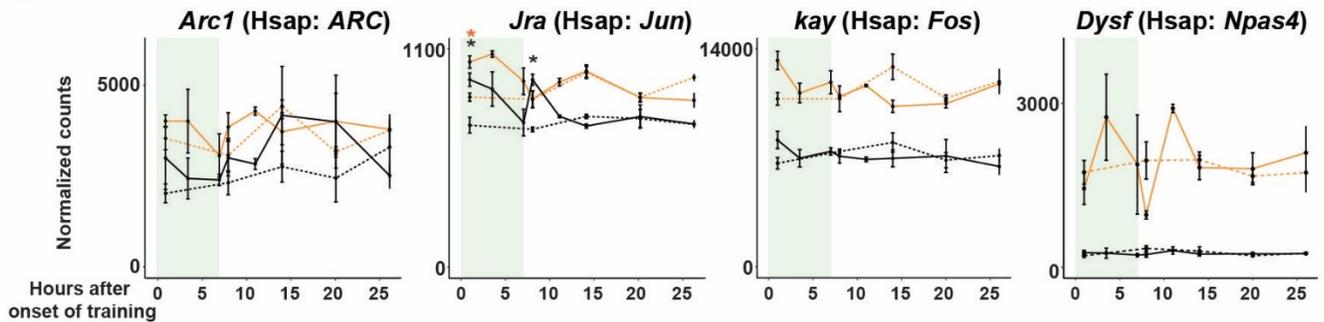
Temporal analysis of TIGs during LTM formation revealed an initial wave of transcription common between the WH and MB at the onset of training, with a secondary induction of transcription occurring at the end of training that is different between the WH and MB (**Figure 3.5**). This dual wave of training induced transcription is in accordance with the current model of neuron-activity regulated transcription, where a common set of IEGs are induced to regulate a downstream wave of more cell-type specific SRG transcription (Mardinly et al., 2016; Sheng et al., 1991; Yap & Greenberg, 2018). To date, no specific IEGs have been identified in *Drosophila* that are known to play a role in long-term memory formation; however, there are candidates based on their responsiveness to a variety of neuronal stimulations (Chen et al., 2016). Here, we looked to identify candidate IEGs which could be regulating the downstream wave of transcription that is specifically induced in the MB at the end of training and early after separation in MB memory tissue.

To identify candidate courtship memory IEGs, 12 known *Drosophila* IEGs, as well as fly orthologs of four well characterized human IEGs (*Arc1/ARC*, *Jra/Jun*, *kay/Fos*, *Dysf/Npas4*), were examined for differential induction across the memory transcriptome time-course (**Figure 3.7A-B**). In the MB, five *Drosophila* IEGs (*Hr38*, *CG14186*, *sr*, *Mctp*, *Cdc7*) and one human IEG ortholog (*Jra*) were found with significant induction. In the WH, eight significantly induced *Drosophila* IEGs (*Hr38*, *CG14186*, *sr*, *Mctp*, *CG8910*, *Cdc7*, *CG14024*, *CG17778*) and one human IEG ortholog (*Jra*) were identified. All significantly increased IEGs in both the WH and MB were found to be induced at one or more of three specific time-points: 1h-DT, 7h-DT, and 1h-AT.

We next looked to select candidate IEGs for further study that could play a direct role in MB training induced transcription contributing to forming stable LTM. Of the six IEGs induced by LTM formation in the MB, three have known functions important for LTM but do not likely impact transcription: *CG14186* (transmembrane protein), *Mctp* (transmembrane protein) and *Cdc7* (protein kinase). Conversely, the other three IEGs – *Hr38*, *sr* and *Jra* – are transcription factors and could play a direct role in downstream training induced transcription.

*Hr38* and *sr* are among the most significantly MB-enriched transcripts (2.98/1.69-fold MB-enriched, respectively) and, unlike *Jra*, are significantly induced at all time-points profiled

throughout training and immediately after separation. Specifically, this study reveals that during LTM formation *Hr38* and *sr* are among the strongest induced transcripts during training in the MB, with maximal expression reached 1h into courtship training (3.63/2.25-fold increase), declining thereafter, but displaying persistent activation up to 1h after separation (2.19/1.63-fold increase), only returning to baseline transcript levels 4h after training ended (**Figure 3.7A**). Taken together, this evidence suggests that the MB-enriched IEG transcription factors *Hr38* and *sr* are top candidates to investigate further for a role in contributing to the memory transcriptional profile induced specifically in the MB during LTM formation.

**A**Previously identified *Drosophila* neuron-activated IEGs**B***Drosophila* orthologs of known mammalian IEGs

## Legend

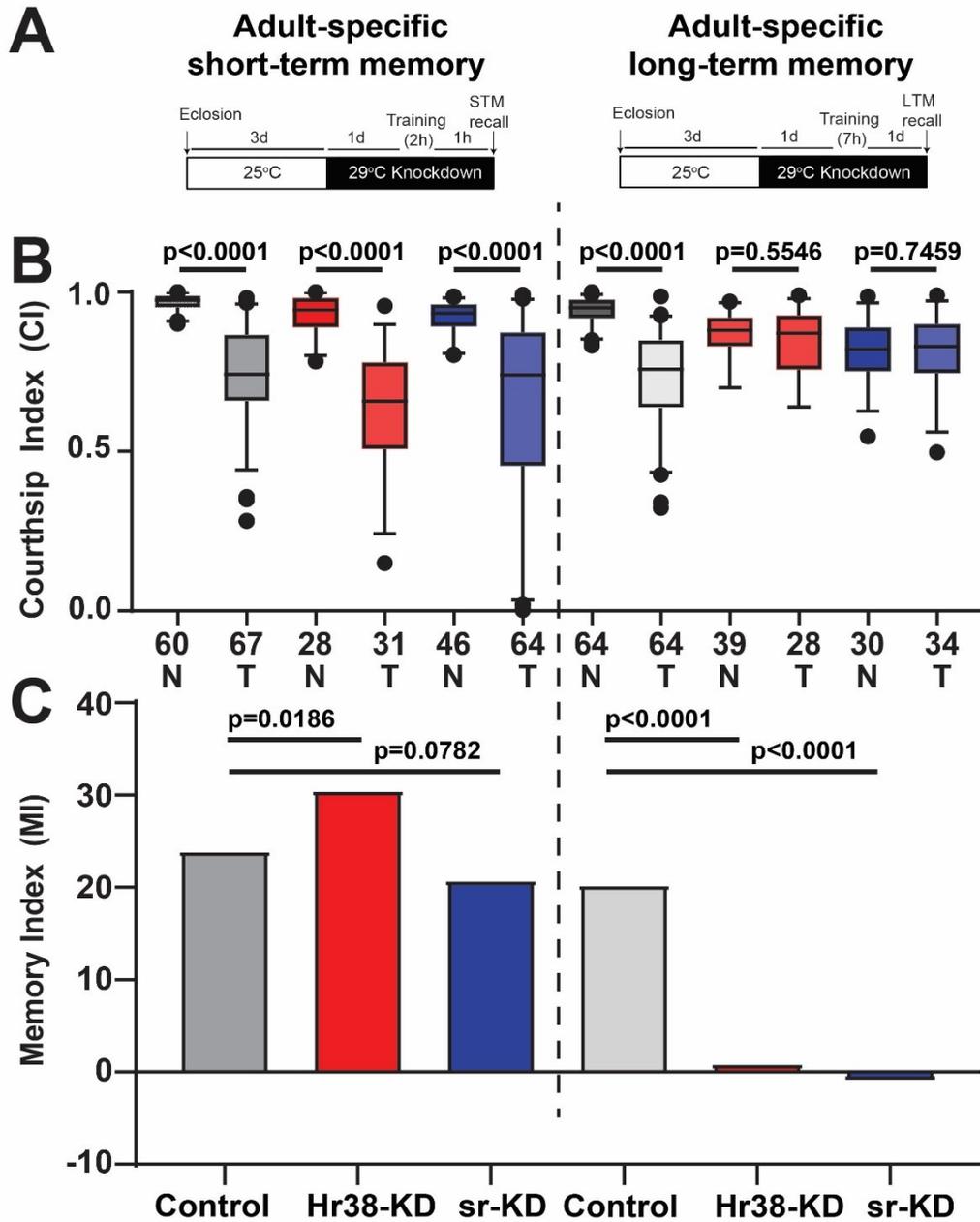


**Figure 3.8 Transcription of known *Drosophila* and fly orthologs of mammalian immediate early genes during LTM formation**

Normalized transcript levels of **(A)** 12 *Drosophila* immediate-early genes (IEGs) previously identified (Chen et al., 2016) and **(B)** four fly orthologs of known mammalian IEGs. Significance identified from differential expression analysis between trained and time-of-day matched naive flies, at time-points noted on the x-axis, is displayed (\*= FDR <0.1).

### 3.7 *Hr38* and *sr* are required in the MB for LTM, but not STM

To provide support that the induction of *Hr38* and *sr* is critical for LTM formation, we looked to determine their requirement for establishing courtship memory independent of development. To do this, temperature sensitive GAL80<sup>ts</sup>-mediated adult-specific knockdown in the MB of either *Hr38* or *sr*, hereafter referred to as Hr38-KD or sr-KD, one day prior to courtship training was performed. Courting was significantly reduced in controls ( $p < 0.0001$ ) following STM training, as well as in Hr38-KD ( $p < 0.0001$ ) and sr-KD ( $p < 0.0001$ ) (**Figure 3.8A, left panel**). STM MI performance was not observed to be significantly reduced in comparison to controls in sr-KD ( $p = 0.0782$ ) but was found to be increased in Hr38-KD ( $p = 0.0186$ ) (**Figure 3.8B, left panel**). With STM able to form following adult-specific knockdown, we then tested to see whether *Hr38* and *sr* had an acute role in LTM. We found that while trained flies in controls had significantly lower courting in comparison to controls ( $p < 0.0001$ ), Hr38-KD ( $p = 0.5546$ ) and sr-KD ( $p = 0.7459$ ) trained flies did not (**Figure 3.8A, right panel**). Accordingly, Hr38-KD and sr-KD both had significantly reduced MI (both  $p < 0.0001$ ) when compared to genetic background controls (**Figure 3.8B, right panel**). Collectively, this data provides evidence that *Hr38* and *sr* are required for LTM formation, but not STM, and suggests that their role as transcription factors play a critical role in regulating downstream training induced genes.



**Figure 3.9 Adult-specific mushroom body knockdowns of IEGs Hr38 and sr in the MB causes defects in LTM but not STM**

(A) Schematic of our approach for achieving adult-specific knockdown in the MB of Hr38 (Hr38-KD) and sr (sr-KD). Adult-specific knockdown was performed using temperature sensitive GAL80<sup>ts</sup> to prevent RNA interference from being active until one-day prior to training. (A) Boxplot distribution of courtship indices (CI) for naive (N) and trained (T) flies of RNAi-mediated

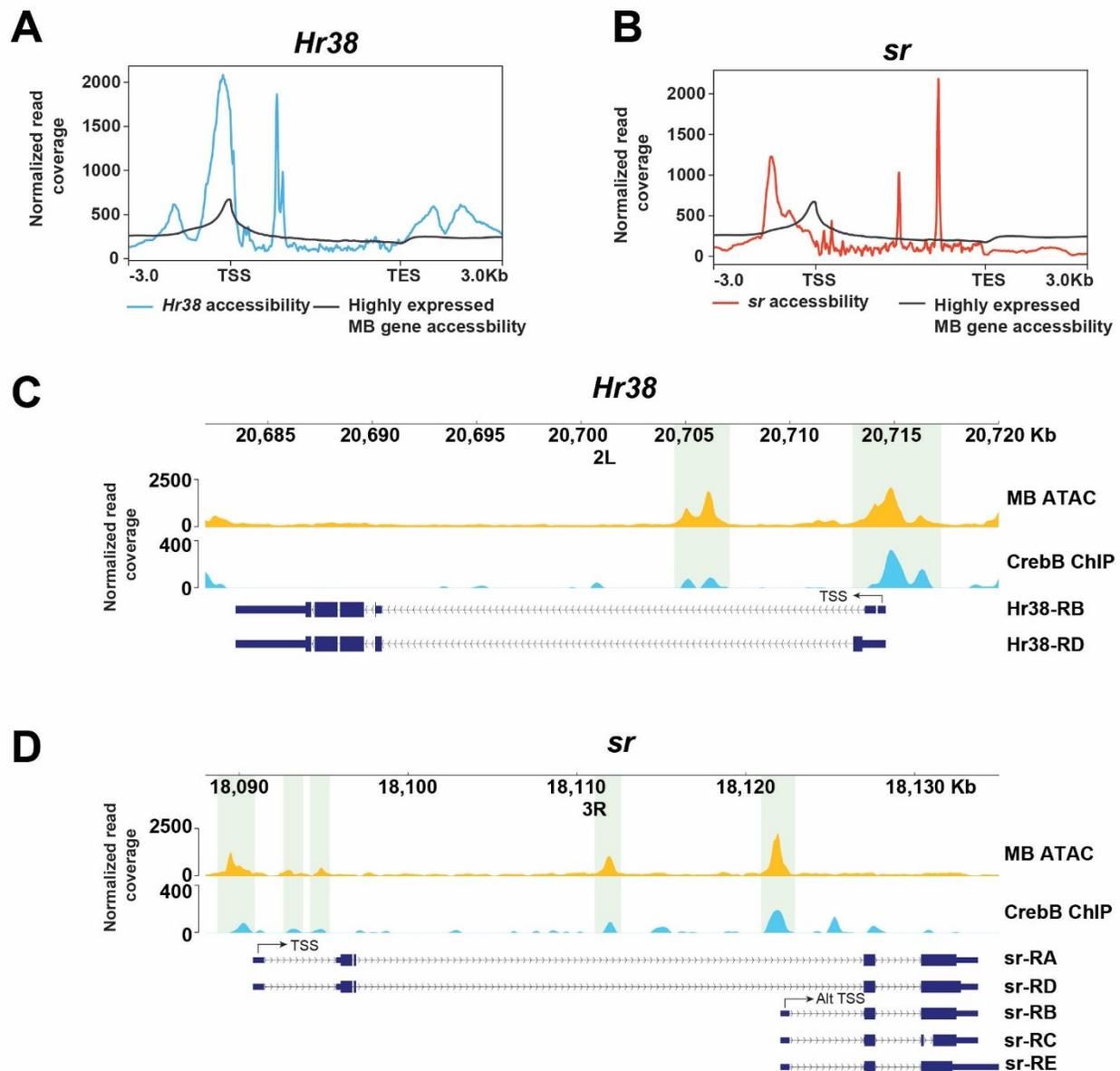
knockdown of Hr38-Kd (red), sr-KD (blue), and genetic background controls (grey) for short-term (STM) and long-term memory (LTM) assays. Statistical comparisons were made between naive and trained flies revealing Hr38 and sr-KD inhibits the formation of specifically of LTM (Mann-Whitney test). Whiskers represent 5-95% percentiles. Total number of flies represented on the row below the x-axis. **(B)** Comparison between control and either Hr38-KD or sr-KD memory indices (MI) show a significant reduction in MI for LTM, but not STM, with exact p-values shown above each assay (randomization test, 10,000 bootstrap replicates).

### 3.8 *Hr38* and *sr* are highly accessible in the MB at regions of CrebB binding

To understand how *Hr38* and *sr* could be activated during LTM formation MB-specific regulatory regions were identified using MB-INTACT ATAC-seq and the presence of transcription factor binding sites based on available ChIP-seq data examined.

First, using MB-specific ATAC-seq, the chromatin landscape of *Hr38* and *sr* in the MB of naive flies at the time corresponding to 1h-DT was observed. *Hr38* and *sr* both displayed regions of high accessibility near, or directly surrounding, their TSS (**Figure 3.9A-B**). For *sr*, notable accessibility was also identified at an alternative TSS roughly 25 kb downstream of the TSS of *sr* isoforms A and D (**Figure 3.9B,D**). When contrasted with highly expressed MB-enriched transcripts (n=545), which was previously identified to validate INTACT specificity, *Hr38* and *sr* are found to be much more accessible, with up to four times as much accessibility around their TSS. This suggests that the naturally accessible chromatin landscape of *Hr38* and *sr* likely plays a role in the immediate and robust induction following neuronal stimulation during courtship training.

CrebB, as a constitutively expressed transcription factor known to regulate IEG expression (Yap & Greenberg, 2018), and which was shown in this study to be required for LTM, was a top candidate to look for potential binding sites located within the *Hr38* or *sr* genes. Here, CrebB binding sites were obtained using publicly available ChIP-seq data from the ENCODE project. CrebB binding was contrasted with MB-specific ATAC-seq chromatin accessibility data, and it was observed that there was prominent CrebB signal within significantly accessible MB chromatin in the TSS of *Hr38*, and in the TSS of all major *sr* isoforms (**Figure 3.9C-D**). The presence of transcription factor binding in an accessible chromatin region near the TSS suggests that the gene could be a direct target. Taken together, these findings suggest that *Hr38* and *sr* could be potentially regulated downstream of CrebB in the MB during courtship training.

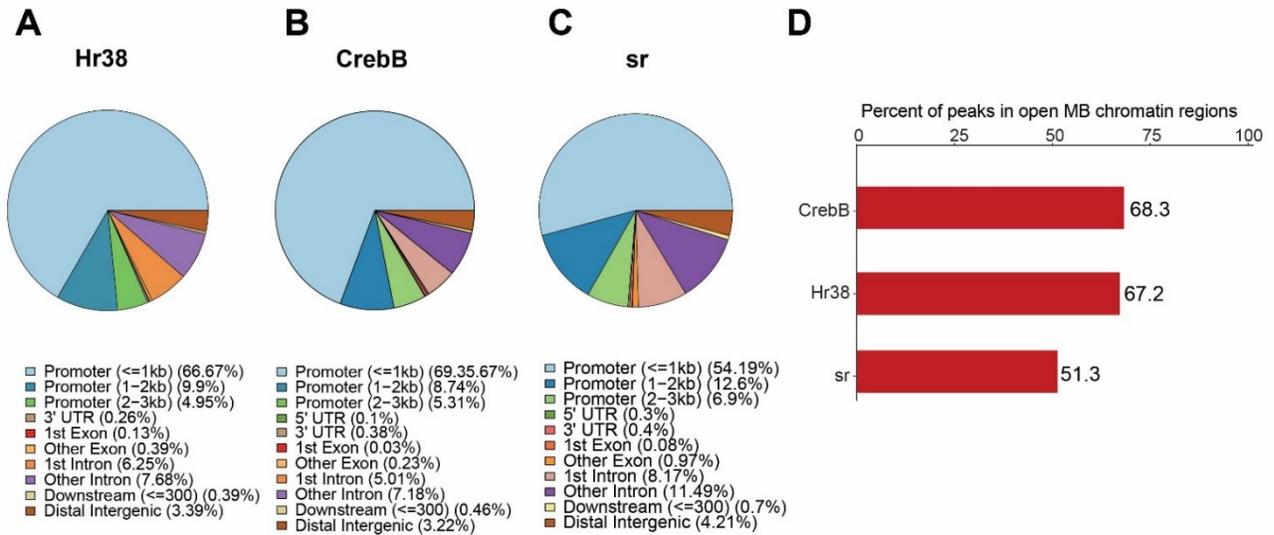


**Figure 3.10 CrebB binds to accessible chromatin regions of *Hr38* and *sr* in the MB**

Bandplot chromatin accessibility profiles of (A) *Hr38* and (B) *sr* shows that these MB memory transcription factor IEGs are far more accessible at their transcriptional start sites in comparison to highly expressed MB-specific genes (n=545). (C) *Hr38* and (D) *sr* genomic tracks including MB-specific ATAC-seq chromatin accessibility information, CrebB ChIP-seq binding site information and *Hr38* and *sr* transcript isoforms. Regions that overlap between CrebB binding signal and significantly accessible MB chromatin are highlighted in light green.

### 3.9 Identifying and characterizing binding sites for Hr38, sr and CrebB

To locate and characterize sites to which Hr38, sr and CrebB bind, publicly available transcription factor ChIP-seq data was obtained from the ENCODE project and annotated to the nearest gene. Overall, 768 Hr38-binding sites that mapped to 689 genes were identified. Hr38-binding sites were highly correlated to the promoter region, with 81.25% of genes being bound within 3 kb of the TSS (**Figure 3.10A**). There were far more genes with binding sites for CrebB and sr than Hr38. Specifically, 3915 CrebB-binding sites that mapped to 2868 genes, and, 5041 sr-binding sites that mapped to 3031 genes, were identified, with 83.4% and 73.69% of binding sites being located within 3 kb of the TSS, respectively (**Figure 3.10B-C**). Binding sites for Hr38, sr or CrebB were then contrasted with regulatory regions of constitutively accessible chromatin in the MB identified using MB-INTACT ATAC-seq. The majority of transcription factor binding sites were found within open chromatin regions in the MB, with 68.3% and 67.2% of CrebB and Hr38-binding sites being significantly accessible, and 51.3% of sr-binding sites (**Figure 3.10D**). These findings suggest that CrebB, Hr38 and sr could play a role in regulating gene transcription in the MB by binding to open chromatin regions.



**Figure 3.11 Annotation and characterization of Hr38, CrebB and sr memory transcription factor binding sites**

Distribution of features annotated peaks for the MB memory transcription factors (A) Hr38, (B) CrebB, (C) sr. (D) Bar graph displaying the overlap of genomic regions with significant binding of Hr38, CrebB or sr with regions of significant chromatin accessibility in the MB, revealing that the majority of memory transcription factor binding sites lie within accessible chromatin regions in the MB.

### 3.10 sr-bound TIGs show a late MB-specific response

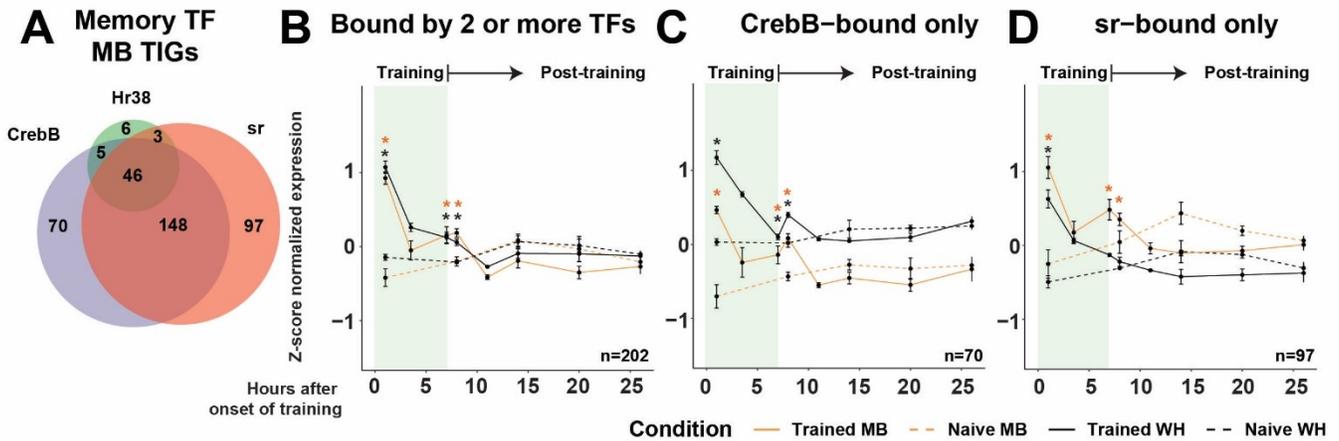
To identify which genes induced as part of the memory-specific transcriptional trace in the MB could be direct downstream targets of Hr38, sr and CrebB, genes with significant transcription factor binding were contrasted with MB TIGs (n=756). In total, of the 756 genes that were induced during LTM formation in the MB, 60, 294, and 269 had binding sites for Hr38, sr and CrebB, respectively. (**Figure 3.11A**). Interestingly, there was a high degree of overlap between transcription factor-bound MB TIGs, with 202 bound by two or more transcription factors and 46 with binding sites for all three (**Figure 3.11A**).

The temporal dynamics of MB TIGs bound by memory transcription factors was examined in three distinct groups: bound by two or more transcription factors (n=202), CrebB-bound only (n=70), or sr-bound only (n=97) (**Figure 3.11A**). WH and MB normalized expression was then plotted for each group of transcription factor-bound TIGs across the transcriptional time-course and significant induction determined.

MB TIGs bound by two or more transcription factors were significantly induced 1h-DT, with transcript levels subsiding by 3.5h-T, followed by significant reactivation of transcription 7h-DT and 1h-AT (**Figure 3.11B**). This was observed in both the MB and WH, with transcript levels similar between both nuclear subsets, and seemingly captures a similar expression trend to what we identified in both MB and WH TIGs (**Figure 3.5B**).

MB TIGs with binding sites only for CrebB also showed significant induction 1h-DT, followed by transcript level reduction by 7h-DT and re-induction 1h-AT in both the WH and MB (**Figure 3.11C**). While these transcriptional trends are similar in both the WH and MB, transcript levels were notably higher in the WH.

Finally, the temporal expression of MB TIGs bound only by sr was observed. Like both other transcription factor-bound memory gene groups, induction was significant 1h-DT in both the WH and MB (**Figure 3.11D**). However, while transcription declined to baseline at 7h-DT and 1h-AT in the WH, transcription was significantly re-induced specifically in the MB at these time-points. Taken together, this provides evidence that sr could play a role, independent of Hr38 or CrebB-binding, in regulating part of the transcriptional trace of memory that is specifically activated in the MB as training progresses.



**Figure 3.12 Temporal analysis of MB training induced genes bound by courtship memory transcription factors during LTM formation**

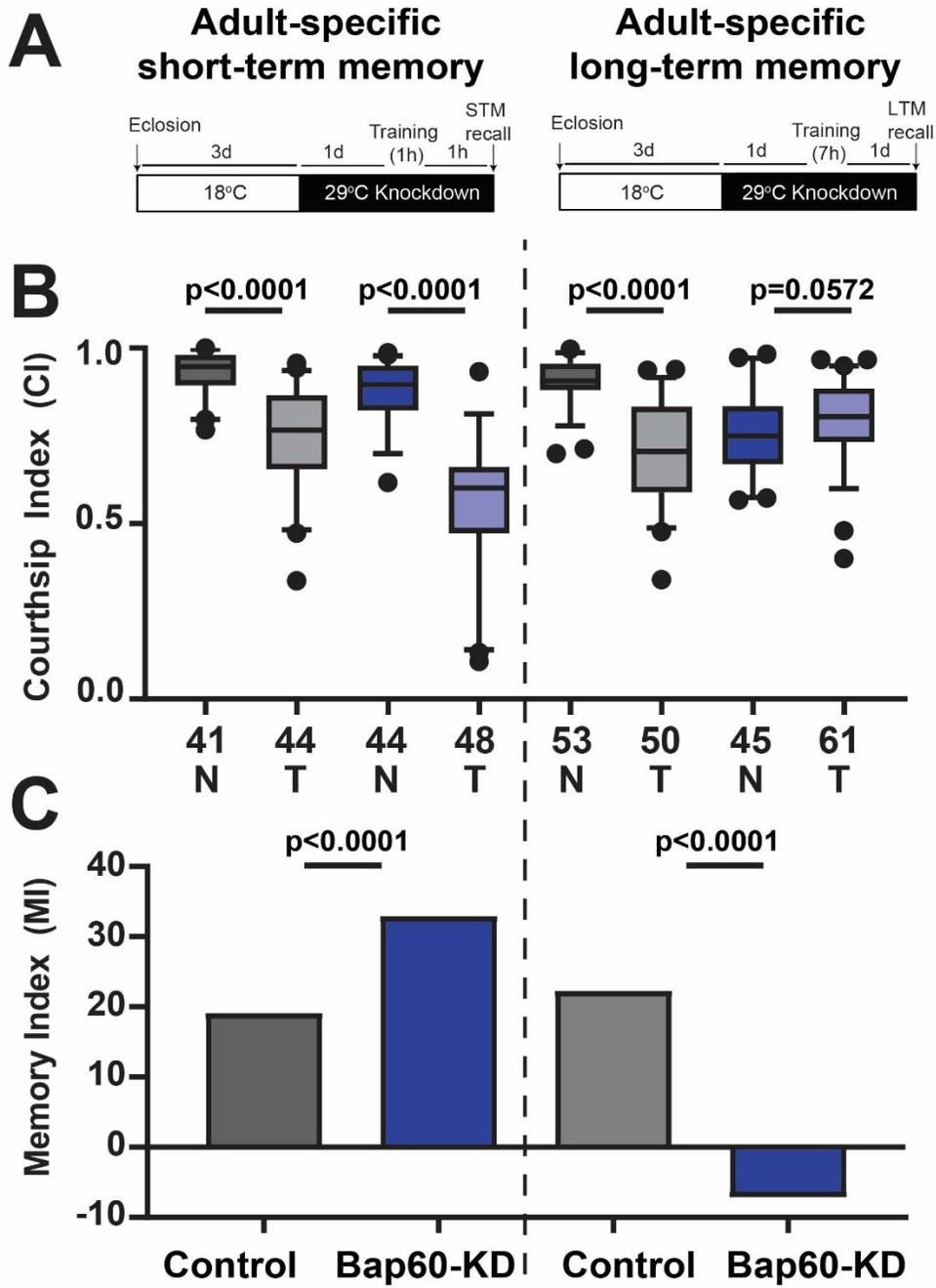
(A) Overlap of binding site overlap of MB memory transcription factors (TF) Hr38, CrebB and sr with training induced genes (TIGs) identified in the MB during LTM formation. Mean z-score normalized expression during LTM formation for (B) MB TIGs with binding sites for two or more memory TF's (C) CrebB-bound only or (D) sr-bound only. Expression was compared between trained, and time-of-day matched naive flies to identify significant induction at time-points noted on x-axis (pairwise t-test, \* =  $p < 0.05$ ).

## Chapter 4: SWI/SNF-mediated chromatin accessibility changes of memory-associated genes during long-term memory formation

Previous work in our lab has studied the role that the SWI/SNF complex plays in post-mitotic MB neurons (Chubak et al., 2019; Nixon et al., 2019). Specifically, a critical role was identified for the SWI/SNF complex in regulating transcription of genes involved in MB neuron remodeling during pupation, as well as MB development in the juvenile adult, ultimately leading to MB morphological defects and deficits in STM and LTM. In this study, we looked to uncover the role that the SWI/SNF complex plays in acute memory processes independent of development by limiting knockdown specifically to the adult fly. In the following experiments, SWI/SNF is shown to be responsible for regulating almost all chromatin accessibility increases and training induced transcription that occurs during LTM formation and that this occurs downstream of the novel courtship memory IEGs *Hr38* and *sr*.

### 4.1 LTM is abolished following adult-specific knockdown of Bap60 in the MB

To assess the role of the SWI/SNF complex in acute memory processes independent of development, we performed an adult-specific knockdown of a core sub-unit of the complex, Bap60. Flies used for courtship memory assays were the product of crossing  $GAL80^{ts}: UAS-Bap60^{RNAi}$  to  $UAS-Unc84::GFP:R14H06-GAL4$ , resulting in flies, hereafter referred to as Bap60-KD, in which we could temporally control the expression of both the *Bap60*-RNAi construct and *UAS-Unc84-GFP* in the MB. Knockdown of *Bap60* and expression of *UAS-Unc84-GFP* was limited to one day prior to STM and LTM courtship training (**Figure 4.1A**). For STM, trained control flies showed significant reduction in courtship activity in comparison to naive flies ( $p < 0.0001$ ), which was also observed in Bap60-KD flies ( $p < 0.0001$ ) (**Figure 4.1B, left panel**). Surprisingly, adult-specific STM MI was found to be significantly enhanced in comparison to controls ( $p < 0.0001$ ) (**Figure 4.1C, left panel**). Courtship behaviour following LTM training was found to be significantly reduced in controls ( $p < 0.0001$ ); however, no change was observed in Bap60-KD males ( $p = 0.0572$ ), resulting in a significant reduction in MI ( $p < 0.0001$ ) (**Figure 4.1 B-C, right panel**). This data supports a critical role for the SWI/SNF complex in establishing LTM, with STM still able to form. As only LTM requires transcription, this suggests that SWI/SNF may play an active role in the regulation of training induced transcription.



**Figure 4.13 Adult-specific knockdown of Bap60 in the MB causes defects in LTM, but not STM**

(A) Schematic of our approach for achieving adult-specific knockdown in the MB of *Bap60* (Bap60-KD). Adult-specific knockdown was performed using temperature sensitive GAL80<sup>ts</sup> to prevent RNA interference from being active until one-day prior to training. GAL80 prevents GAL4-mediated transcription at 18 °C but becomes inactivated at 29 °C, allowing for the expression of Bap60-KD. (B) Boxplot distribution of courtship indices (CI) for naive (N) and trained (T) flies of RNAi-mediated knockdown of Bap60-KD (blue) and its genetic background control (mCherry-KD, grey) for short-term (STM) and long-term memory (LTM) assays. Statistical comparisons were made between naive and trained flies revealing that Bap60-KD inhibits the formation specifically of LTM (Mann-Whitney test). Whiskers represent 5-95% percentiles. Total number of flies represented on the row below the x-axis (C) Comparison between control and Bap60-KD memory indices (MI) show a significant reduction in MI for LTM, and a significant increase for STM, with exact p-values shown above each assay (randomization test, 10,000 bootstrap replicates).

## 4.2 Bap60-KD suppresses transcription and chromatin accessibility post-training

The SWI/SNF complex plays an integral role in promoting transcriptional activation by generating open chromatin at promoters and enhancers through ATP-dependent nucleosome sliding, ejection or restructuring (López & Wood, 2015; Vogel-Ciernia et al., 2013b). The role of SWI/SNF in regulating transcription is conserved across species with much learned through work on cancer and in a developmental context (Mashtalir et al., 2018; Son & Crabtree, 2014). Recent work has indicated SWI/SNF is also responsive to environmental signals, being recruited by IEGs and brought to enhancer regions to establish accessible chromatin (Vierbuchen et al., 2017). To date this SWI/SNF transcriptional regulatory mechanism has only been displayed *in vitro*, with much still to be learned about how SWI/SNF responds to environmental signals *in vivo*.

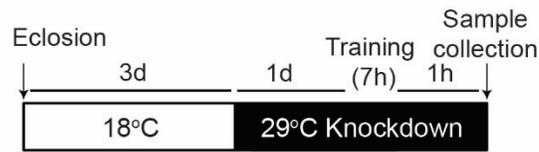
With our findings that Bap60-KD specifically abolishes LTM, and with STM still able to form, we hypothesized that loss of Bap60 could be impacting the ability for SWI/SNF to make chromatin more accessible for genes induced during LTM formation. To assess the role that SWI/SNF plays in LTM transcriptional regulation we used INTACT to isolate MB nuclei from Bap60-KD and mCherry-KD control flies one-hour after LTM training had finished and contrasted with time-of-day matched naive male flies (**Figure 4.2**). This directly corresponds to the critical window of MB-specific training induced transcription identified in **Chapter 3**.

Transcript level and chromatin accessibility information was determined by performing RNA-seq and ATAC-seq on INTACT-isolated MB nuclei from the same biological sample. After processing RNA-seq libraries, 8626 highly expressed genes were used for differential expression analysis where contrasts were made between all possible pairwise comparisons between conditions: 1) trained controls vs naive controls; 2) trained Bap60-KD vs naive Bap60-KD; 3) naive Bap60-KD vs naive control; 4) trained Bap60-KD vs trained control. Genes were considered differentially expressed if they had an FDR <0.05. (**Figure 4.3 A-D**).

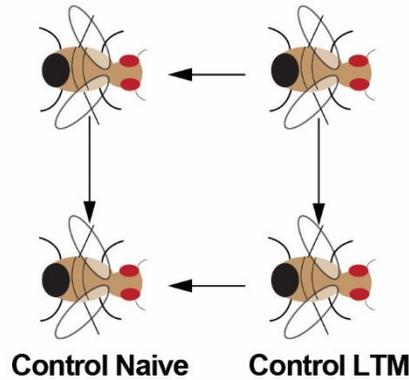
Knockdown of Bap60 had a major impact on training induced transcription, with only 17 genes induced after training in comparison to 128 that were induced in mCherry-KD controls, with 8 genes consistent between both (**Figure 4.3 A-B**). Comparing naive Bap60-KD to mCherry-KD flies, there were 129 genes with higher transcript levels in controls compared to Bap60-KD before training (**Figure 4.3 C**). There were more differences observed in transcript levels between Bap60-KD and control flies after training, with 387 genes found to have significantly higher transcript

levels in controls compared to Bap60 after training (**Figure 4.3 D**). Collectively, these results suggest an important role for Bap60 in regulating transcription after training.

To determine which genes had dynamically regulated chromatin accessibility changes during LTM formation we performed differential peak analysis on 16293 regulatory regions using the R package DiffBind (**Chapter 2.11.2**). In controls we identified 690 regions with significantly increased accessibility in trained vs naive flies (FDR<0.05), corresponding to 593 genes (**Figure 4.4 A**). In Bap60-KD flies we only identified one gene with increased accessibility after training (**Figure 4.4 B**). Between Bap60-KD and control flies, there were only 14 genes with less chromatin accessibility in Bap60-KD flies compared to controls before training (**Figure 4.4 C**). The effect of Bap60-KD on chromatin accessibility was remarkable after training, with Bap60-KD flies found to have 1571 regions with less accessibility after training, representing 1221 genes, when compared to controls after training (**Figure 4.4 D**). Taken together with our RNA-seq findings, these results suggest that normal Bap60 expression is required for almost all inducible transcription and increased chromatin accessibility changes that occur in the MB after LTM training.

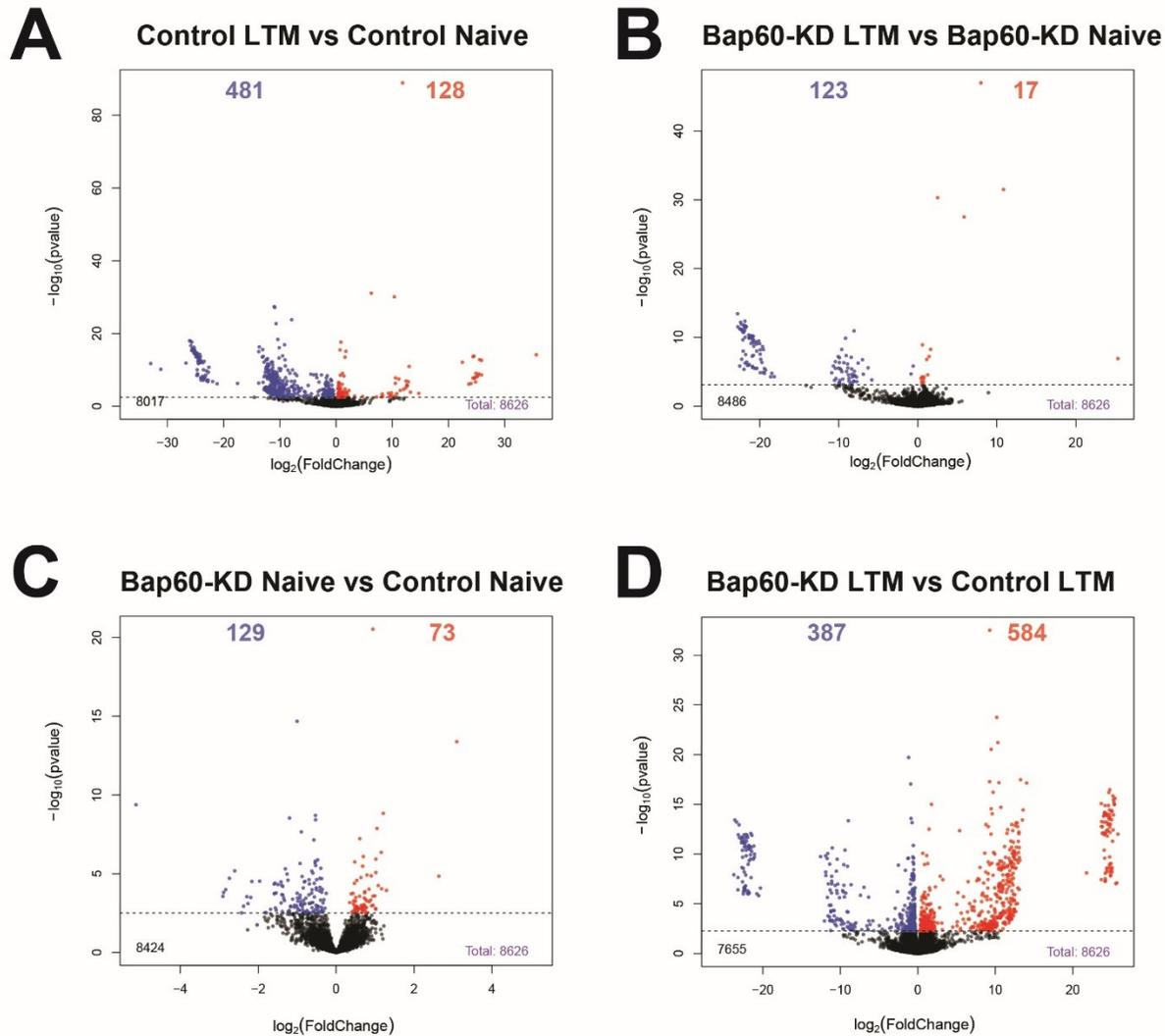


**Bap60-KD Naive    Bap60-KD LTM**



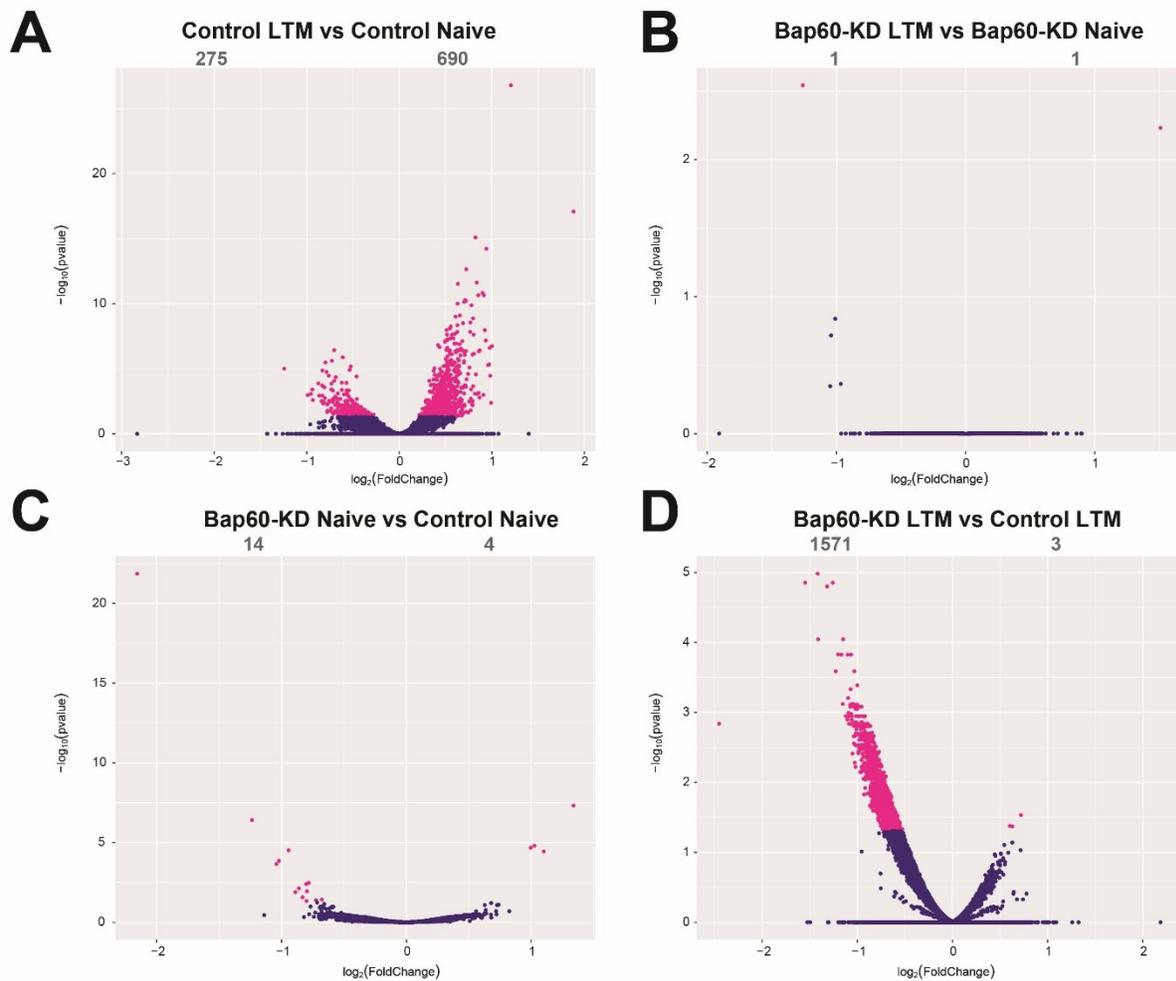
**Figure 4.14 Approach for profiling the gene regulatory role of SWI/SNF during LTM formation**

To profile the role that SWI/SNF plays in regulating training induced expression during LTM, we performed INTACT on Bap60-KD and control flies one hour after LTM training had concluded, as well as on time-matched naive flies. INTACT bound nuclei were processed to generate RNA-seq libraries, to profile transcript level changes, as well as ATAC-seq libraries, to profile chromatin accessibility changes, both from the same biological sample. Arrows represent contrasts made for differential expression and differential accessibility analysis: 1) trained controls vs naive controls; 2) trained Bap60-KD vs naive Bap60-KD; 3) naive Bap60-KD vs naive control; 4) trained Bap60-KD vs trained control.



**Figure 4.15 Identifying Bap60-dependent memory-regulated transcript level changes**

Differential expression analysis results visualized as a volcano plot for INTACT-obtained mushroom body (MB) transcript level changes between **(A)** control LTM vs control naive **(B)** Bap60-KD LTM vs Bap60-KD naive **(C)** Bap60-KD naive vs control naive **(D)** Bap60-KD LTM vs control LTM. 8626 genes were used for differential expression analysis and declared significant if  $FDR < 0.05$ . Training induced genes are highlighted in red, and memory repressed genes are highlighted in blue.



**Figure 4.16 Identifying Bap60-dependent memory-regulated chromatin accessibility changes**

Differential accessibility analysis results visualized as a volcano plot for INTACT-obtained mushroom body (MB) transcript level changes between **(A)** control LTM vs control naive **(B)** Bap60-KD LTM vs Bap60-KD naive **(C)** Bap60-KD naive vs control naive **(D)** Bap60-KD LTM vs control LTM. 16293 peaks were used for differential accessibility analysis and declared significant if  $FDR < 0.05$  (highlighted in pink). Peaks with increased accessibility are located on the right side of each plot, and peaks with decreased accessibility are located on the left side.

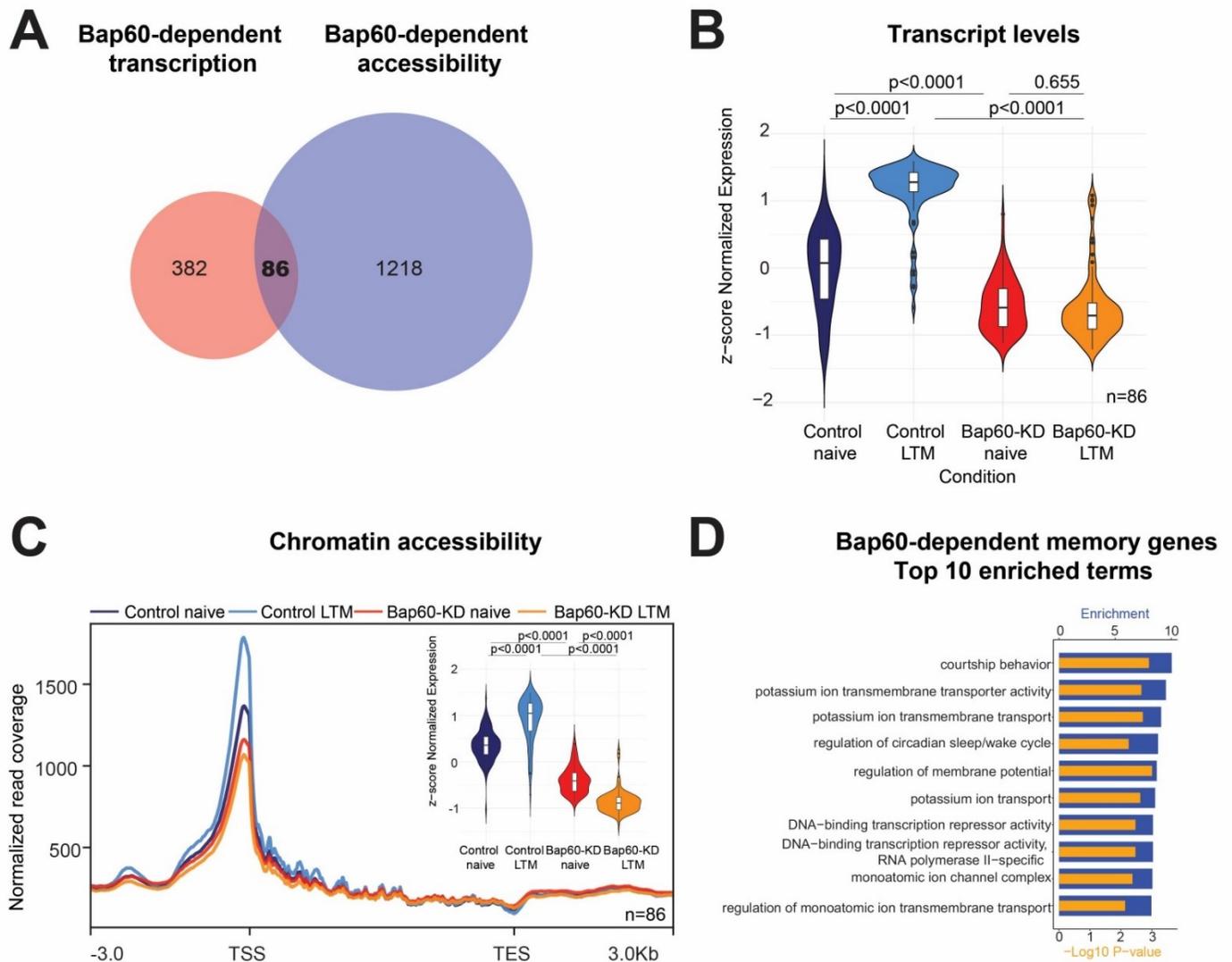
### 4.3 Bap60-dependent memory genes play integral roles in LTM formation

Bap60-KD dramatically reduced overall memory gene inducibility. We next looked to identify genes that required Bap60 for both transcript induction and chromatin accessibility increases. These Bap60-dependent changes could represent genes where memory stimulated SWI/SNF chromatin accessibility changes could be directly impacting gene transcription required for LTM formation.

Overall, we identified 86 genes that required Bap60 for increased transcription and chromatin accessibility after training (**Figure 4.5A**) (**Table 4.1**). These Bap60-dependent memory genes are significantly induced in controls ( $p < 0.0001$ ), with this inducibility lost in Bap60-KD ( $p = 0.655$ ). Interestingly, the baseline expression of these genes was also impacted with a significant decrease in transcript levels of Bap60-dependent genes in Bap60-KD compared to controls ( $p < 0.0001$ ) (**Figure 4.5B**). Similar trends were also observed for chromatin accessibility changes in Bap60-dependent genes. Bap60-dependent memory genes indicated a significant increase in accessibility in controls ( $p < 0.0001$ ) which was not observed in Bap60-KD, which instead showed a significant decrease in accessibility ( $p < 0.0001$ ) (**Figure 4.5C**). Baseline naive fly TSS accessibility was also significantly decreased in Bap60-KD when compared to controls ( $p < 0.0001$ ). Taken together, we have identified a set of training induced genes which require Bap60 for increased TSS accessibility which could be contributing to loss of their memory-sensitive transcript induction during LTM formation

We next performed GO analysis on Bap60-dependent memory genes to functionally characterize the roles that these genes could be playing during LTM formation. The most enriched term we identified among Bap60-dependent genes was “courtship behavior” (10.03-fold enriched) which was annotated to the genes *pros*, *shep*, *Sh*, *CASK*, *dsf* and *fne* (**Figure 4.5D**). Other significantly enriched terms were relevant to memory-associated processes including “regulation of membrane potential”, “synaptic signaling”, “calcium ion transmembrane transport”, “axon development”, “neuron projection” and “regulation of gene”. We also noted enrichment of the term “sleep” and “circadian sleep/wake cycle” which we previously showed was enriched among MB-only TIGs (**Figure 3.6**). Collectively, this data provides evidence that Bap60 plays a critical role in regulating the chromatin accessibility and training induced transcription of a set of genes

that play important biological roles relevant to memory in the post-training period during LTM formation.



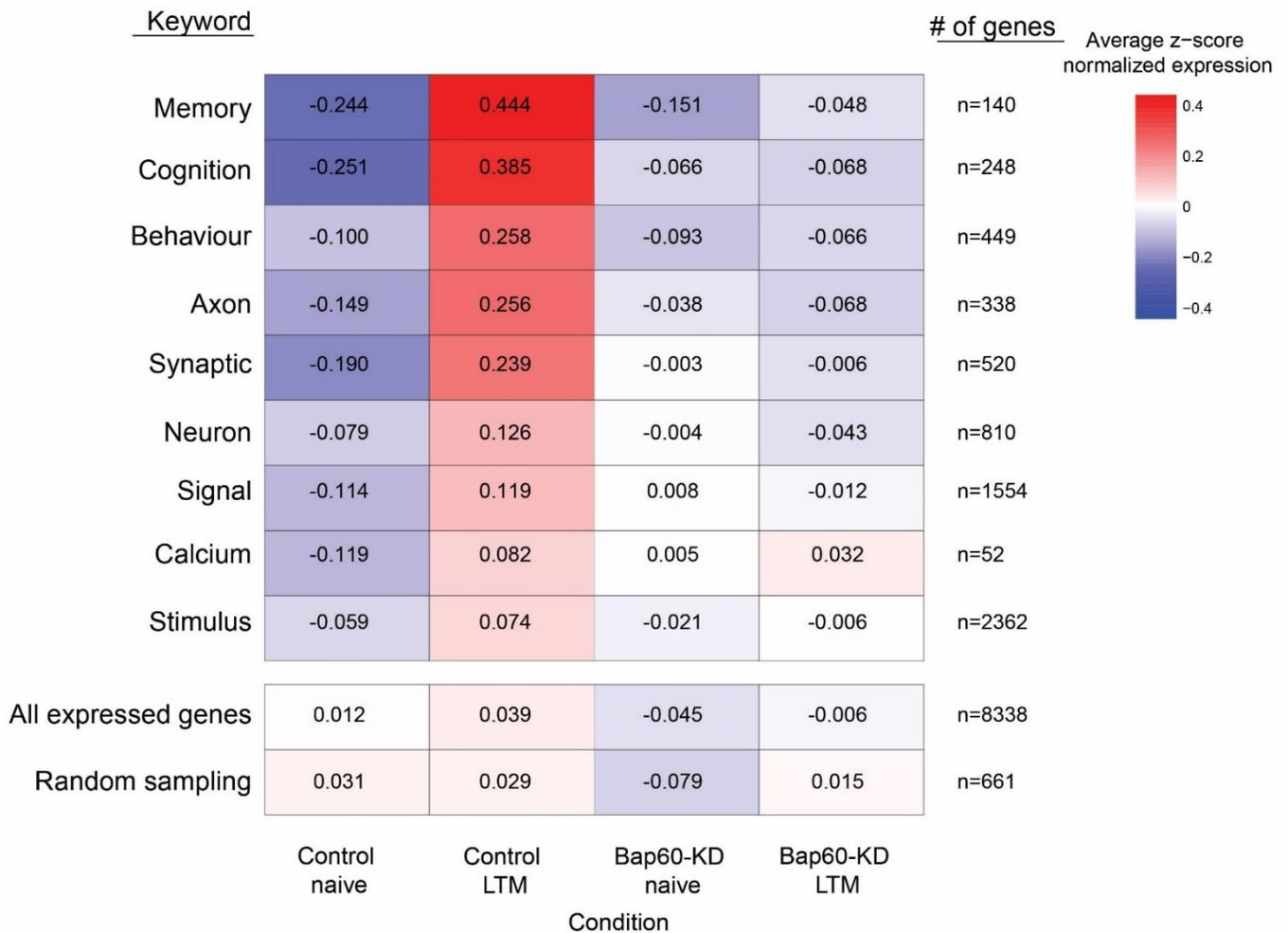
**Figure 4.17 Bap60 regulates both the transcription and chromatin accessibility increases of a subset of memory-relevant genes**

(A) Genes that require Bap60 for normal LTM induced transcription ( $n=468$ ) were overlapped with genes that require Bap60 for normal LTM induced chromatin accessibility increases ( $n=1304$ ), to reveal 86 Bap60-dependent memory genes. (B) Violin plot distribution of z-score normalized transcript levels for each profiled condition for Bap60-dependent memory genes. Significant expression changes between groups were determined using a pairwise t-test, with p-values noted between contrasted conditions. (C) Bandplot chromatin accessibility profiles for Bap60-dependent

memory genes, plotted by condition with mean normalized read coverage spread across the gene body shown. Significant accessibility changes were determined using a pairwise t-test between groups using reads that directly overlapped the TSS of Bap60-dependent memory genes. P-values are noted above between contrasted conditions. **(D)** Top 10 enriched terms identified by gene ontology (GO) analysis for Bap60-dependent memory genes. Significant GO terms (FDR <0.05) were identified for biological processes, molecular functions, and cellular components by contrasting Bap60-dependent memory genes to a background of expressed MB genes identified in INTACT-isolated nuclei.

#### 4.4 Bap60 regulates general transcript inducibility of memory-relevant genes during LTM

Bap60-dependent memory genes were found to be enriched for terms associated with gene expression, including at least 17 genes with known transcription factor DNA-binding properties including *pros*, *dac*, *bsh*, *dsf*, *Lim1*, *mamo*, *Glut4EF*, and *jim*. By regulating the expression level of these transcription factors, Bap60 might have broad ranging impacts on training induced transcription. To test this, we analyzed expression levels of all genes annotated to GO terms that included important memory-associated keywords ranging from highly specific “memory” (n=140) and “cognition” (n=248) to more general keywords like “signal” (n=1554) and “stimulus” (n=2362) (**Figure 4.6**). We noted that the transcriptional induction that was observed in controls following training was relative to the specificity of the keyword in relation to memory. This normal LTM induced transcription was dramatically reduced in Bap60-KD, if not eliminated completely. For example, genes annotated to terms containing the keyword “memory” (n=140) had a z-score normalized expression increase of 0.6881 in controls but only 0.10 in Bap60-KD, an ~85% difference. Complete elimination of memory-sensitive inducible expression was prominently noted for genes annotated to terms containing the keywords “axon”, “cognition”, “synaptic” and “neuron”. Even more general keywords like “stimulus” (n=2362) which had a modest z-score normalized expression increase of 0.13251 in controls, was reduced to only 0.02773 in Bap60-KD, an 80% decrease. To establish a baseline a random set of genes was analyzed. These showed almost no change between naive and trained in controls, with only slight induction seen in Bap60-KD. Everything considered, these results strengthen the argument that Bap60 is required for the inducible expression of a wide network of genes during LTM formation and include those that are directly Bap60-dependent and additional genes that are regulated downstream of Bap60-dependent memory genes.



**Figure 4.18 Bap60 directly and indirectly regulates gene inducibility during LTM formation**

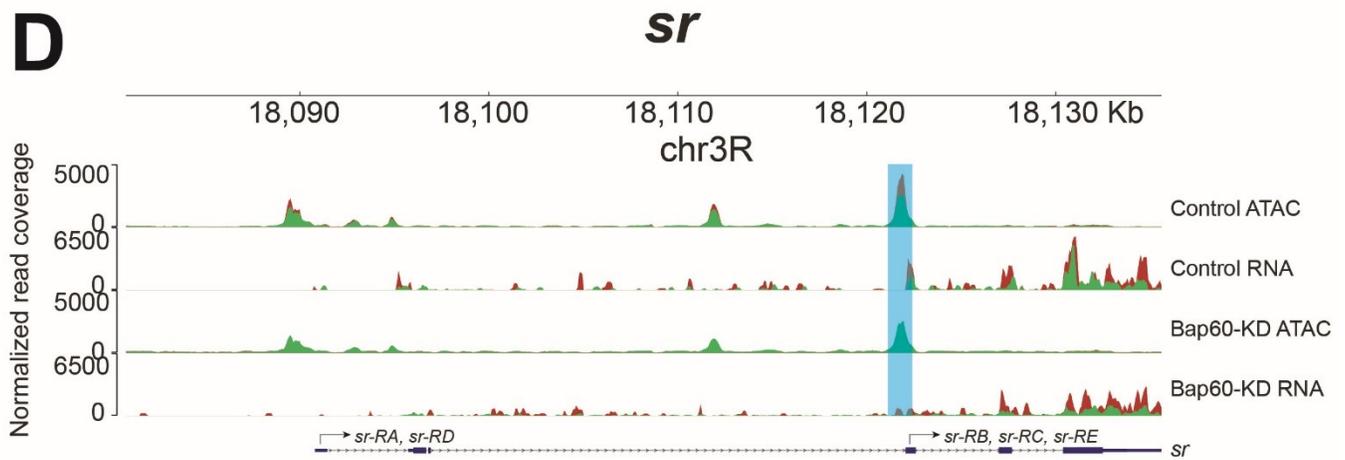
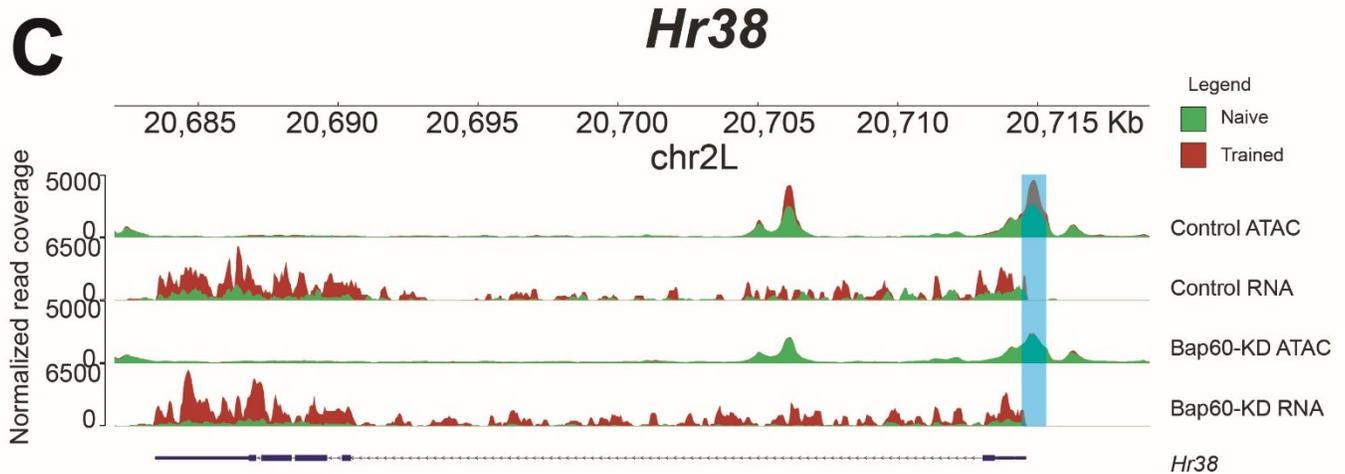
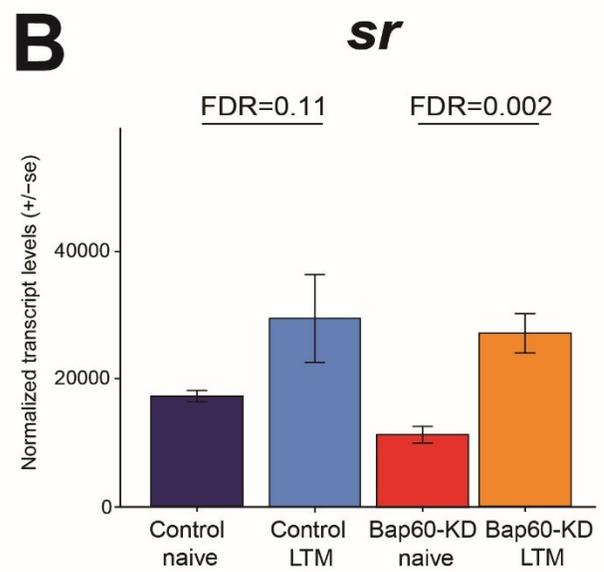
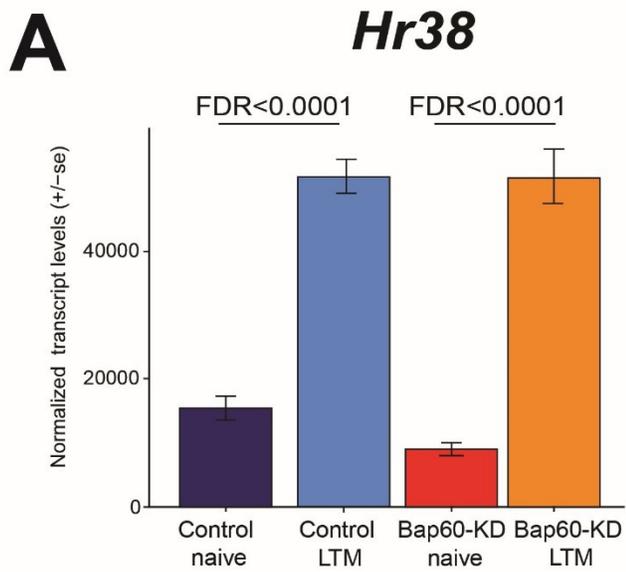
Memory-relevant keywords were used to extract genes annotated to gene ontology terms containing the keyword. Average z-score normalized expression was calculated for each gene annotated to the keyword and the mean calculated and contrasted across all four conditions. We also performed this comparison for all expressed genes, as well as for a random set of 661 expressed genes, representing approximately 5% of genes with expressed proteins in *Drosophila*, as well as all expressed genes. Mean z-score normalized expression was calculated for each set of genes and results displayed as a heatmap.

#### 4.5 Bap60-KD does not impact transcription of *Hr38* and *sr* in the MB

Our results suggest that upon knockdown of Bap60 a large network of genes is impaired from becoming transcribed during LTM formation. We then asked where in the hierarchy of LTM transcriptional induction does Bap60, and in extension SWI/SNF, enact its chromatin-regulating role. Specifically, if loss of Bap60 impairs general gene inducibility – is IEG expression also limited? Here, we looked to contextualize our work with our previous findings that the transcription factors *Hr38* and *sr* are courtship memory IEGs required specifically in the MB for LTM formation, but not STM, by assessing their transcriptional induction and chromatin accessibility in Bap60-KD compared to controls.

Upon Bap60-KD almost all training induced transcription is eliminated, however, we identified 17 DEGs (FDR<0.05) that were still induced in Bap60-KD post-LTM training. Eight genes were also identified as DEGs in mCherry-KD controls and included: the lipoproteins *Apoltp* and *Jeb*, which we have consistently identified as induced during LTM, the associative learning required Dopamine/Ecdysteroid receptor *DopEcR*, as well as the non-transcription factor insect IEG *CG14186*. Most notably, however, both *Hr38* and *sr* were found to still have significantly robust induction in Bap60-KD following training with fold-changes of 5.70 and 2.38, respectively, and were not significantly different from transcript levels in controls after training (p=0.99 and p=0.93, respectively) (**Figure 4.7A-B**).

Both *Hr38* and *sr* were found to have significantly increased chromatin accessibility in mCherry-KD flies after training. For *Hr38*, this increased chromatin accessibility was significant around the TSS of both major isoforms (FDR<0.0001) (**Figure 4.7C**). Conversely, *sr* had a significant increase in the accessibility at one TSS associated with 3 of the 5 major *sr* isoforms (*sr-RB*, *sr-RC*, *sr-RE*; p=0.014) (**Figure 4.7D**). These three *sr* isoforms also had greater RNA-seq read coverage, implying they may play a larger role in LTM than the two isoforms transcribed from the *sr* TSS upstream. These increases in *Hr38* and *sr* TSS chromatin accessibility were not observed in Bap60-KD flies after training. Taken together, these results indicate that Bap60 plays a role in mediating chromatin accessibility changes in the MB of TSS regions for *Hr38* and *sr* but that this does not impact training induced transcription. In the context of the hierarchy of LTM-induced transcription, this suggests that SWI/SNF plays a critical role in regulating downstream SRG memory transcription, with IEG inducibility left intact.



**Figure 4.19 Bap60 regulates memory-gene transcription downstream of Hr38 and sr**

**(A-B)** Normalized transcript levels of courtship memory IEGs **(A)** *Hr38* and **(B)** *sr* for Bap60-KD and controls after LTM training compared to naive flies. Error bars display standard error. FDR values are shown between trained and naive conditions to indicate significant training induced transcription. **(C-D)** Track figures showing RNA and ATAC-seq normalized read coverage for **(C)** *Hr38* and **(D)** *sr*. Trained (red) and naive (green) conditions are overlapped for both Bap60-KD and controls. Regions with a significant increase in chromatin accessibility (FDR<0.05) in controls that is not observed in Bap60-KD are highlighted in blue. TSS for different *sr*-isoforms are indicated below the tracks.

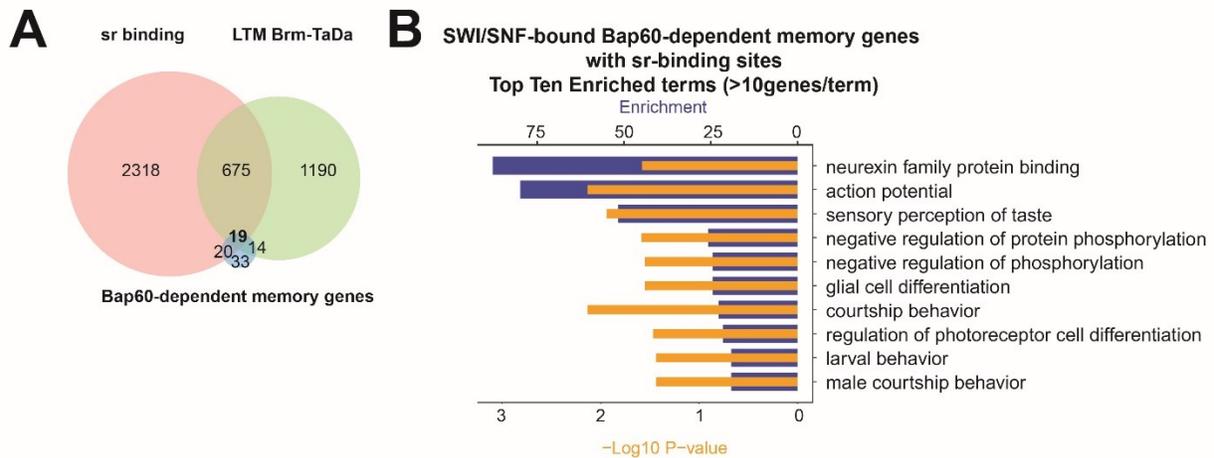
#### 4.6 SWI/SNF and *sr* bind to a subset of memory-relevant Bap60-dependent genes

We next looked to identify which of the Bap60-dependent memory genes that we identified were directly bound by SWI/SNF during LTM formation using Targeted DamID (TaDa) data generated by our collaborator, Dr. Francisco Martin-Castro (Autonomous University of Madrid). We identified 1898 genes that were bound by *brm*, the ATPase associated with the SWI/SNF complex, in the MB during LTM formation. MB specific *brm* binding sites were found at 33 of 86 Bap60-dependent memory genes (**Figure 4.8A, Table 4.1**). These 33 genes likely represent direct SWI/SNF targets during LTM formation.

Our results show that Bap60-KD impacts memory SRG, but not IEG, inducibility in the MB early after training has ended. Contextualized with our transcriptome data, which suggests that memory-sensitive genes bound by the courtship IEG *sr* are part of the MB-specific transcriptional program that becomes activated as training progresses, we wondered if there was a connection between *sr* and Bap60-dependent memory genes bound by SWI/SNF. 19 of the 33 Bap60-dependent memory genes also had *sr* binding sites (**Figure 4.8A, Table 4.1**). These 19 genes were enriched for many of the same functional groups that were identified among our full list of 86 Bap60-dependent memory genes (**Figure 4.8D**). These included memory-relevant terms like “neuroxin family protein binding” (*btsz, CASK*, 87-fold enriched), “action potential” (*sh, Ca- $\alpha$ IT*, 79-fold enriched), “courtship behavior” (*sh, CASK, pros*, 22.7-fold enriched) and “dendrite development” (*tutl, pros, jim*, 8.33-fold enriched).

Finally, out the 19 candidate *sr*-bound Bap60-dependent SWI/SNF target memory genes, four were found with *sr* and *brm* binding within 1kb of each other. These are genes where *sr* and *brm* could work together to regulate the transcription of the candidate memory gene during LTM. Included among the four genes was the transmembrane protein *tutl*, which is involved in axonal pathfinding during development (Al-Anzi & Wyman, 2009), and has six SWI/SNF binding-sites and two *sr*-binding sites. Of note, *sr* is found to bind directly on the *tutl* TSS, but also to a region slightly downstream of the TSS, likely an enhancer region, which directly overlaps with SWI/SNF binding (**Figure 4.9A**). The zinc finger transcription factor encoded by *jim*, involved in dendrite morphogenesis and chromatin silencing (Iyer et al., 2013), is found to have two regions co-bound by *sr* and *brm*, including one directly over the TSS (**Figure 4.9B**). The homeobox transcription factor *pros*, with a well-established role in development and shown to have a role in courtship

behavior (Grosjean et al., 2004, 2007), is found to have four prominent sr-binding sites, including one located downstream of the TSS in a possible enhancer region which is proximal to SWI/SNF binding (**Figure 4.9C**). *Sty*, which is thought to play a role in negatively regulating several signaling pathways (Reich et al., 1999), has seven sr-binding sites, including one in a potential enhancer region downstream of the TSS which overlaps with SWI/SNF-binding (**Figure 4.9D**). In all four genes (*tutl*, *pros*, *jim* and *sty*), the TSS shows significantly increased chromatin accessibility that is lost following Bap60-KD. Altogether, these four genes present themselves as top candidates for further study and may provide valuable insight into the mechanisms of how SWI/SNF is recruited to chromatin.

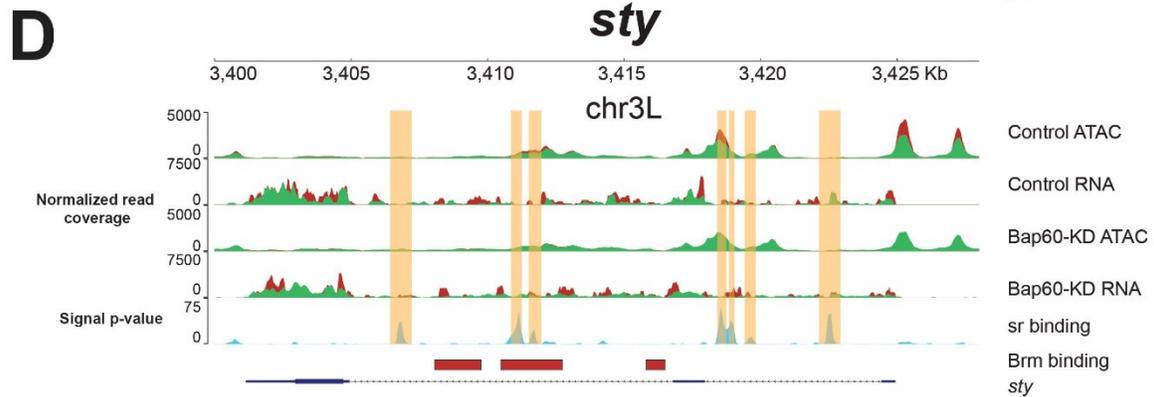
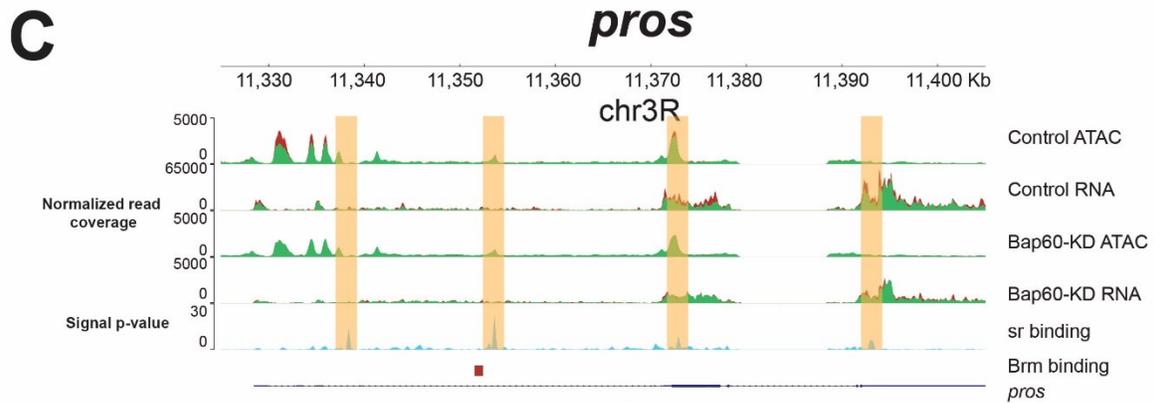
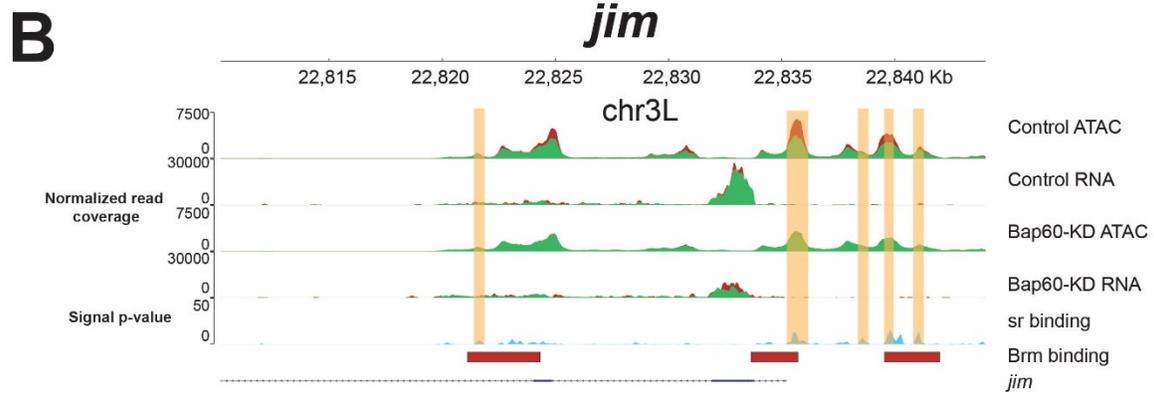
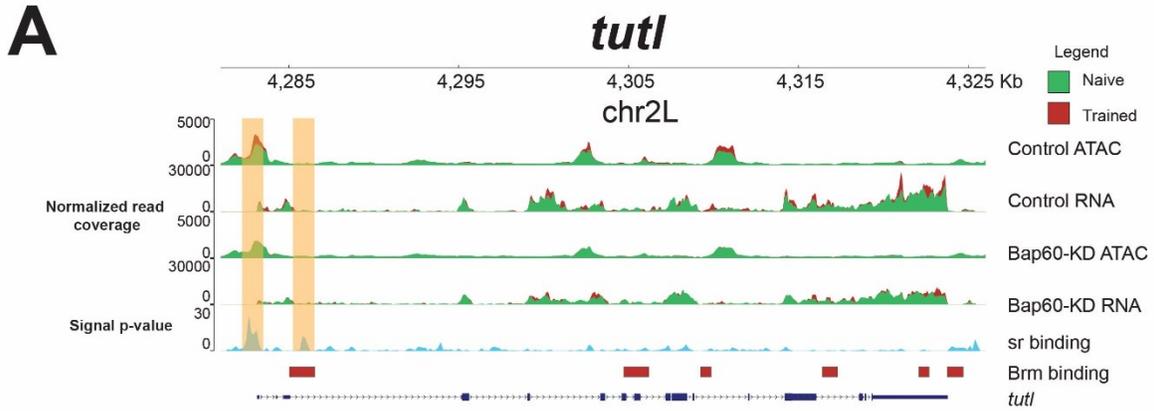


**Figure 4.20 Identifying and characterizing SWI/SNF and sr-binding sites for Bap60-dependent memory genes**

(A) Overlap of genes with sr-binding sites identified by ChIP-seq, brm-TaDa bound genes during LTM formation and Bap60-dependent memory genes. (B) Top ten enriched terms identified by GO analysis of SWI/SNF-bound Bap60-dependent memory genes with sr-binding sites contrasted to a background of highly expressed MB genes.

**Table 4.7 Bap60-dependent memory genes and presence of SWI/SNF and sr binding sites**

Gene name	sr bound	SWI/SNF bound	Gene name	sr bound	SWI/SNF bound
Ace	No	Yes	HGTX	No	No
AOX4	No	No	hig	No	No
AstA-R2	No	No	Hk	No	No
beat-Ia	No	No	IA-2	Yes	No
beat-IIa	No	Yes	Imp	Yes	No
bru3	Yes	No	ImpE2	No	No
bsh	No	No	jeb	Yes	No
<b>btsz</b>	<b>Yes</b>	<b>Yes</b>	<b>jim</b>	<b>Yes</b>	<b>Yes</b>
<b>Bx</b>	<b>Yes</b>	<b>Yes</b>	kek2	No	Yes
<b>Ca-alpha1T</b>	<b>Yes</b>	<b>Yes</b>	kek3	No	Yes
CadN2	No	Yes	Lim1	No	No
<b>CASK</b>	<b>Yes</b>	<b>Yes</b>	<b>Liprin-gamma</b>	<b>Yes</b>	<b>Yes</b>
Cda9	No	No	<b>mamo</b>	<b>Yes</b>	<b>Yes</b>
CG10936	Yes	No	Mid1	Yes	No
CG12071	No	No	<b>mim</b>	<b>Yes</b>	<b>Yes</b>
CG12484	No	Yes	mirr	Yes	No
CG12605	No	No	msn	Yes	No
CG18265	No	No	Nckx30C	No	Yes
CG34354	No	No	<b>pros</b>	<b>Yes</b>	<b>Yes</b>
CG42342	No	Yes	Rab2	No	No
CG42795	No	Yes	RabX1	No	No
CG44838	Yes	No	<b>Rbp6</b>	<b>Yes</b>	<b>Yes</b>
CG45002	Yes	No	Rdl	No	Yes
CG4577	Yes	No	robo1	Yes	No
CG6329	No	No	Sap47	Yes	No
<b>CG8398</b>	<b>Yes</b>	<b>Yes</b>	scrt	No	Yes
<b>chinmo</b>	<b>Yes</b>	<b>Yes</b>	<b>Sh</b>	<b>Yes</b>	<b>Yes</b>
Cnx99A	Yes	No	Shab	No	No
cpx	No	No	shakB	No	No
CR44334	No	No	shep	Yes	No
CR45436	No	No	<b>sm</b>	<b>Yes</b>	<b>Yes</b>
CR45706	No	No	smal	Yes	No
ctp	No	No	SRPK	Yes	No
dac	Yes	No	<b>sty</b>	<b>Yes</b>	<b>Yes</b>
dimm	No	No	Task6	No	No
dpr15	No	Yes	<b>Tet</b>	<b>Yes</b>	<b>Yes</b>
Dscam2	No	Yes	tio	Yes	No
dsf	No	No	Tlk	Yes	No
esg	No	No	<b>tna</b>	<b>Yes</b>	<b>Yes</b>
fne	No	No	trv	No	No
Frq1	No	No	<b>tutl</b>	<b>Yes</b>	<b>Yes</b>
GluClalpha	No	Yes	Ubc6	No	No
<b>Glut4EF</b>	<b>Yes</b>	<b>Yes</b>	v	No	No



**Figure 4.21 SWI/SNF and sr have proximal or overlapping binding sites for a subset of Bap60-dependent memory genes**

Track figures for four genes that are candidates for co-localized binding of SWI/SNF and sr during LTM formation. **(A)** *tutl* **(B)** *jim* **(C)** *pros* **(D)** *sty*. Tracks show overlap of trained (red) and naive (green) RNA and ATAC-seq normalized read coverage, brm-TaDa binding post-LTM training and ChIP-seq identified sr-binding sites. Regions with significant sr-binding are highlighted in yellow.

## Chapter 5: Discussion

The work presented in this thesis provides valuable new insight into the mechanisms underlying long-term courtship memory formation. Using cell-type specific methods, we extensively profile the memory transcriptome of MB memory-forming neurons of *Drosophila*. Our results reveal a critical MB-specific memory transcriptional program that becomes activated as training progresses, which in part may be regulated by the novel courtship memory IEGs *Hr38* and *sr*, which we show are required specifically for LTM. Further, using MB-specific RNA-seq and ATAC-seq, we provide the first evidence that the SWI/SNF complex regulates memory gene inducibility during acute memory processes.

### 5.1 The transcriptional trace of memory becomes activated in the MB towards the end of training

In this study, we thoroughly characterized training induced changes in transcription in both MB memory-specific and WH memory non-specific tissue during LTM formation. This work provides significant advances in our understanding of the transcriptional trace of memory by generating and profiling the most extensive RNA-seq time-course of *Drosophila* memory formation to date. Specifically, using the courtship conditioning memory assay, we profiled training induced transcript level changes at three time-points during the courtship training period, as well as at several novel time-points after training had ended (**Figure 3.5A**). This approach revealed that most transcriptional activity during LTM formation occurs in both memory-specific and memory non-specific tissue during the courtship training period (**Figure 3.5B**). In fact, like what we found previously, there are very few genes with significantly changed transcript levels one day after training, which correlates to when LTM recall is normally tested (Jones et al., 2018; McBride et al., 1999). This suggests that courtship memory does not broadly induce lasting changes to transcription and other mechanisms for maintaining and storing a persistent LTM should be investigated. Potential mechanisms to explore include long-lasting changes to neuronal epigenetic marks in the MB. In mammals, epigenetic marks have been shown to be rapidly changed during learning and that these changes can be stably maintained up until 30 days later (Halder et al., 2016; Miller et al., 2010; Narkaj et al., 2018; Zovkic, 2021). Early findings in flies suggest a similar mechanism, with training inducing persistent changes to epigenetic machinery in the MB for other forms of memory (Hirano et al., 2016).

By contrasting TIGs identified in the MB or the WH, we were able to identify the transcriptional trace of memory that is common or distinct between memory-specific or non-specific tissue (**Figure 3.6A**). We identified 423 TIGs common between both the MB and WH. These TIGs shared a consistent transcriptional trend, becoming strongly induced early after training onset, becoming reactivated at the end of training, and then returning to baseline thereafter, and may reflect the transcriptional trace of courtship behaviour (**Figure 3.6B**). Upon exposure to a female fly, the male fly brain undergoes transcriptional changes almost immediately as it displays its stereotypical courtship behaviour (Barajas-Azpeleta et al., 2018; Fujita et al., 2013; Takayanagi-kiya & Kiya, 2019). These transcriptional changes, which do not induce a lasting LTM and are not specific to the MB, are measurable up to one-hour after a two-hour courtship training period, and may be reflected in what we observe here (Barajas-Azpeleta et al., 2018).

We also identified 815 TIGs that were only induced in the WH during LTM formation (**Figure 3.6A**). These genes were not induced early into courtship training, but instead showed robust induction specifically in the WH as courtship training progressed (**Figure 3.6B**). Perhaps most interestingly, in both the MB and WH, a cyclical pattern in transcription was observed in naive flies which followed the incubator day/night cycle. We suggest that training disrupts a circadian-linked transcriptional program at the end of courtship training that is expressed in the whole fly brain. Further, with recent evidence pointing to the MB playing a non-canonical role in circadian regulation of gene expression, our dataset could be used to ask new questions to help progress this fascinating new field of research (Machado Almeida et al., 2021).

We identified 333 TIGs that were only induced in the MB during LTM formation (**Figure 3.6A**). These genes were similarly induced in both the WH and MB early into courtship training, however, they were only re-induced in the MB towards the end of training and early into the courtship rest period, when memory consolidation is thought to occur (**Figure 3.6D**). This suggests that MB-only TIGs may reflect the transcriptional trace specific and required for memory formation, a notion supported by the induction of genes known already to be involved in courtship memory, including those associated with cAMP-mediated signaling (**Figure 3.7**). Indeed, the work presented in this thesis highlights the period at the end of courtship training and early after training as the critical window for LTM-required transcription. This can also be correlated with our findings of the translational requirements for courtship memory. Specifically, we showed that translation is

required during LTM formation, but is dispensable during training (**Figure 3.1**). This suggests a model where a MB-specific memory transcriptional program becomes activated as training progresses, which is then translated after training to form proteins that contribute to the formation of a long-lasting memory. Further study, consisting of protein level validation and a courtship memory screen for select candidate MB-only TIGs will be needed to provide evidence of this, however, the proposed timing fits with the current model of courtship memory consolidation. Specifically, the importance of the early post-training period for memory consolidation has been highlighted through study on the translation regulator cytoplasmic polyadenylation element-binding protein Orb2, and post-learning sleep requirements (Dag et al., 2019; Keleman et al., 2012; Krüttner et al., 2015).

Functionally, our results emphasize the dynamic regulation of metabolic pathways throughout LTM formation starting from the beginning of training in both the MB and WH, with distinct metabolic requirements activated in the MB as training progresses, including the induction of tricarboxylic acid cycle components (**Figure 3.7**). The results are not surprising as the brain is one of the most energetically demanding organs, requiring metabolic fine tuning to meet energy demands (Sgammeglia & Sprecher, 2022). In fact, memory consolidation requires a consistent energy supply, with starved flies unable to form LTM (Plaçais & Preat, 2013). The metabolism of memory is currently of great interest to neuroscientists, and we are only beginning to understand the importance that energy homeostasis in both neurons and supporting tissue, like glia, play in LTM formation (de Tredern et al., 2021; Plaçais et al., 2017; Silva et al., 2022). By profiling both memory-specific MB neurons, as well as WH tissue, this work offers a wealth of information that can be used to open new lines of inquiry into the dynamics of metabolism during courtship memory formation.

## **5.2 Hr38 and sr play a critical role as IEGs in *Drosophila* courtship memory**

This study demonstrates that profiling the training induced transcriptome in the MB is an effective way to identify target genes to test for a role in memory. By contrasting our data with known IEGs in *Drosophila* and mammals, we were able to select and identify a role for *Hr38* and *sr* specifically in forming LTM, with a dispensable role in STM (**Figure 3.9**). While *Hr38* and *sr* have been previously shown to be induced during LTM formation, as well as in response to early courting behaviour, this work greatly expands our understanding of how these novel courtship

IEGs are transcribed during LTM (Fujita et al., 2013; Jones et al., 2018; Raun, 2019; Takayanagi & Kiya, 2019). Specifically, we show that *Hr38* and *sr* are strongly induced immediately as courtship training begins, with high induction throughout training that is sustained up until one hour after training has ended (**Figure 3.8A**). The continued expression *Hr38* and *sr* throughout training is surprising and runs in contrast to what is generally understood about IEGs, which are thought to be induced quickly but then subside. Currently, we can only speculate as to why courtship IEGs are expressed continuously throughout training, but there are at least two potential biological explanations. First, unlike other memory assays where IEGs have been identified, courtship memory is one continuous training session, with males naturally spacing their courting attempts. This could cause intermittent activation of memory neurons and re-induction of courtship IEGs. Second, *Hr38* (~31 kb) and *sr* (~11 kb) are longer than the average fly gene, and significantly larger than mammalian IEGs like *c-fos* (~3.4 kb) and *c-jun* (~3.2 kb) (Chen et al., 2016), likely impacting the rate of translation, and thus not displaying the same transcriptional trends observed in mammals. Further study should focus on examining protein level changes of *Hr38* and *sr* during LTM formation to elucidate this finding.

This study provides insight into the chromatin landscape underlying *Hr38* and *sr*. Specifically, we show that the transcriptional start sites of *Hr38* and *sr* are highly accessible in the MB, which is consistent with the finding that the transcriptional start sites of fly ARGs are more accessible in general (**Figure 3.11A-B**) (Chen et al., 2016). CrebB, which we show is required for LTM, was found to have binding within the accessible transcriptional start sites of *Hr38* and *sr* in the MB, suggesting that they are downstream targets for CREB binding during LTM formation (**Figure 3.2; Figure 3.11C-D**). Finally, using transcription factor ChIP-seq binding profiles for *Hr38*, *sr* and CrebB, we show that MB TIGs that are only bound by *sr* contribute to the MB-specific memory transcriptional trace that becomes activated as courtship training progresses. Together, to strengthen our conclusions that 1) *Hr38* and *sr* are downstream targets of CREB-mediated transcription during LTM and 2) *sr* regulates the MB-specific response during memory, further study could be performed that focus on profiling the memory transcriptome following either CrebB or *sr*, and *Hr38*-knockdown.

### 5.3 The SWI/SNF complex is required for memory gene inducibility in the MB during LTM formation

This study showed that adult-specific knockdown of Bap60, a core sub-unit of the SWI/SNF complex, results in the defective formation of LTM, with STM still able to form (**Figure 4.1**). Following adult-specific knockdown of Bap60, we used RNA-seq and ATAC-seq to profile transcript level and chromatin accessibility changes one-hour after courtship training had ended, which correlates with the window of MB-specific memory transcription we previously identified. Strikingly, we show that Bap60 knockdown causes the near complete loss of training induced chromatin accessibility changes and inducible gene transcription (**Figure 4.3-4.4**). We show that loss of Bap60 impacts memory transcription downstream of IEGs, with SWI/SNF seemingly redundant for the expression of *Hr38* and *sr* (**Figure 4.7**). Using Brm-TaDa data provided by a collaborator, we identified direct targets of the SWI/SNF complex during LTM and in combination with binding sites for *sr*, we identified 19 memory-relevant genes that offer top candidates for further study (**Figure 4.8 -4.9, Table 4.1**). Many of these candidate genes are transcription factors, and we show that Bap60 is required for the inducible expression of a wide network of genes during LTM both directly and regulated downstream (**Figure 4.6**). Together, this is the first study to provide insight into the role of the SWI/SNF complex during acute memory processes, and with a well-studied role in neurodevelopmental disorders, may have important implications for understanding processes crucial to cognition in health and disease.

What remains to be determined is how knockdown of Bap60 prevents the SWI/SNF complex from regulating memory-gene inducibility. In flies, it has been shown that the SWI/SNF complex is still able to form properly without Bap60, even though it is a core component (Mashtalir et al., 2018). Recent evidence has shown that a close Bap60 ortholog in mice, SMARCD2, is required for the recruitment of transcription factors to the promoters of granulocyte differentiation genes (Priam et al., 2017). This finding is especially interesting in the context of other work that has shown that the SWI/SNF complex is recruited to enhancers by the IEG transcription factor AP-1 complex to regulate an activity-dependent transcriptional program in mice fibroblasts (Vierbuchen et al., 2017; Yap & Greenberg, 2018). Taken together with our finding that SWI/SNF is not required for IEG induction, this suggests a model where Bap60 may interact with IEGs, which we speculate could include *sr*, which then guide the SWI/SNF complex to memory genes where it makes the required chromatin accessibility changes for training induced transcription. To

test this hypothesis, recent genetic tools that have been developed for use in *Drosophila* make it possible to identify protein interactors that have previously been difficult to capture through traditional mass spectrometry – immunoprecipitation approaches. Specifically, advances in TurboID, where a lasting biotin label is placed on proximal proteins, makes it possible to capture more transient protein interactors. This approach has been recently successful for the histone demethylase KDM5 (Yheskel et al., 2023), which suggests that the use of TurboID in the context of Bap60 or other SWI/SNF components could provide important mechanistic insight into how the SWI/SNF complex regulates memory gene inducibility.

#### **5.4 Research limitations**

To study the molecular mechanisms underlying memory in *Drosophila*, we modified and optimized the INTACT method to isolate MB nuclei for memory-tissue specific transcriptomic and epigenomic profiling. This approach has been highly effective in providing novel insight into memory neurons and how LTM is formed within them, as exemplified by this thesis and other work from our lab (Jones et al., 2018; Nixon, 2020; Nixon et al., 2019; Raun, 2019). Given the low number of cells that are likely activated during memory formation, the argument could be made that profiling the entirety of MB nuclei is becoming an approach as antiquated as bulk RNA-seq in whole tissue or micro-array. The argument for finer spatial resolution of the memory transcriptome is strengthened by recent studies which have elucidated a tripartite loop of neurons synapsing in the MB  $\gamma$  lobe which are required for formation and recall of cVA-retrievable courtship memory (Dag et al., 2019; Keleman et al., 2012; Krüttner et al., 2015; X. Zhao et al., 2018) – why not profile them directly? Currently, there remains large technical and practical limitations to investigating such a small set of memory-required cells. Single cell RNA-seq technology, while advancing rapidly, is still in its infancy. In fact, due to sequencing depth limitations, experiments using this technology are only able to capture changes in the highest expressed transcripts and due to cost limitations, temporal analysis is often not explored (Dal Molin & Di Camillo, 2019; Gil-Martí et al., 2022). Additionally, the bioinformatic tools to properly and reproducibly analyze single cell RNA-seq data are still in development (Gibson, 2022). Further, while the cell-type specific tools exist in *Drosophila* to allow for the expression of UAS-Unc84::GFP in very small sets of neurons, the number of fly heads that would need to be processed to extract enough material for INTACT RNA-seq would not be feasible. Conversely, MB INTACT

allows for cost-effective, deep-sequencing of memory-specific neurons that is reproducible. In fact, while this work is the first to test and identify *Hr38* and *sr* for a novel role in memory, our previous transcriptome work had also identified these courtship IEG as TIGs one hour after courtship training, among other consistently induced genes including (but not limited to) *jeb*, *svr*, *CG44247*, *CG33204*, *CG12531*, *CG14186*, *Apoltp* (Jones et al., 2018; Raun, 2019). This emphasizes the high quality of this data. Taken together, while MB INTACT captures nuclei that may not contribute directly to the memory transcriptome, our approach to broadly identify training induced transcriptional changes in the MB, followed by tissue-specific knockdown and further genetic dissection, is a valuable way to better elucidate the mechanisms underlying memory formation.

A major driving force behind the ENCODE (Encyclopedia of DNA elements) project is to systematically map regions of transcription, transcription factor association, chromatin structure, and histone modification – and importantly – make this data publicly available for researcher use (Abascal et al., 2020; ENCODE Project Consortium, 2012). Here, we make use of multiple ENCODE datasets to identify candidate genes that the memory transcription factors *Hr38*, *sr* and *CrebB* bind to during LTM formation. These transcription factor ChIP-seq data sets were generated from fly prepupa or embryo's, which limits the biological conclusions we can make from this binding site information as it is not specific to memory tissue and the signals that drive transcription factor binding during development are fundamentally different from what can be expected during LTM formation. One way that we tried to mitigate this limitation is by looking for transcription factor binding within accessible regions of chromatin in the MB, which is an established approach to correlate binding in non-profiled tissues and avoid costly transcription factor ChIP-seq assays (Schmidt et al., 2017). This is especially relevant to our findings where we showed that *CrebB* binds to accessible chromatin regions of *Hr38* and *sr* in the MB (**Figure 3.11**). While transcription factor ChIP-seq assays are likely not possible using the MB INTACT approach due to limited material and a lack of good antibodies, exciting new genomics tools have opened the possibility to accurately profile transcription factor binding in the small cell populations. Specifically, the sister technologies CUT&Tag and CUT&RUN, have enabled cell inputs as low as 5,000 cells to accurately profile histone PTMs, and chromatin-associated proteins. Importantly, the work flow for these protocols is compatible with INTACT and have been shown to effectively profile the chromatin landscape of *Drosophila* photoreceptors and neural stem cells (Henikoff et

al., 2020; Jauregui-Lozano et al., 2021; Keegan et al., 2023). Optimizing these techniques with our MB-specific GAL4 driver will be a critical next step if we are to use these tools to profile binding of our transcription factors of interest in the MB during LTM formation.

Currently, while we use the training induced transcriptome to identify a role for the IEGs *Hr38* and *sr* in regulating long-term courtship memory formation, this work does not functionally validate the SRGs that are activated downstream. Further, while we have highlighted some top candidate genes directly regulated by SWI/SNF during LTM formation, they too remain to be functionally validated for a role in memory (**Figure 4.9**). Thus, some aspects of this work remain exploratory, and largely hypothesis generating, requiring additional follow-up. Using the GAL80<sup>ts</sup> system to functionally validate candidate genes for a role in memory, like what we did for *Hr38*, *sr* and *Bap60*, is an effective approach but is limited by a lower throughput using the courtship memory assay due to flies developing slower at 18°C. With 333 potential MB-only TIGs or 97 MB TIGs with *sr* binding sites, tools like CUT&RUN/TAG may be useful to help refine and select top candidate target genes for functional validation.

## 5.5 Conclusions, future directions, and research implications

To summarize, this study presents the most thorough investigation of the *Drosophila* memory transcriptome to date and identifies a critical timeframe during courtship training when memory transcription is induced specifically in memory tissue. Like other memory assays, we show that translation is required for long-term courtship memory (T Tully et al., 1994), suggesting that these training induced transcripts are processed into proteins that play a functional role in establishing a persistent long-lasting memory.

Using the advances made by this work, we now have novel insight into the hierarchy of transcription factors that are activated in the MB during LTM formation and can begin to build a model (**Figure 5.1**). Specifically, during courtship training, female rejection signals which act upon dopaminergic neurons, and pheromone signals which stimulate calcium signalling (Raun et al., 2021), converge upon the cAMP pathway to activate the constitutively expressed transcription factor CrebB, which we show is required for LTM. CrebB then likely binds to highly accessible regions of *Hr38* and *sr* in the MB, to robustly induce their transcription early into courtship training. CrebB, *Hr38* and *sr* then likely bind to and increase the transcription of genes with roles

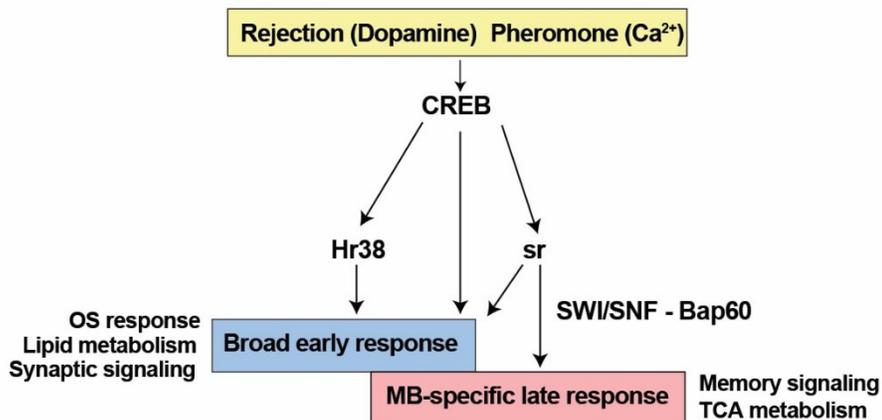
in lipid metabolism, synaptic signaling and the response to oxidative stress. The increase in transcription of these genes is observed in both memory-specific and non-specific tissue and represents a broad early transcriptional response likely caused by general courting behaviour and interaction with the female fly. However, as courtship training progresses, a MB-specific memory transcriptional program is activated which is likely mediated in part by binding of sr. These genes have known roles in memory, including components of the cAMP pathway and dopamine signalling, but also reflect the unique metabolic requirements of MB memory tissue – a current area of focus in the field.

The hierarchy of courtship memory transcription factors that we establish from the training induced transcriptome (**Chapter 3**) fits with our work that presents the first data implicating the SWI/SNF complex in directly regulating the inducibility of memory genes during LTM formation (**Chapter 4**). Here, SWI/SNF acts downstream of the IEGs Hr38 and sr to increase chromatin accessibility and gene transcription during the early post-training period, which coincides with the timing of the activation of the MB-specific memory transcriptional program we identify. Further, we show that among direct SWI/SNF-regulated memory genes are also overlapping targets of sr – suggesting that the potential to act together to regulate memory gene transcription, however the exact mechanism remains to be explored.

Collectively, the work in this thesis thoroughly characterizes the transcriptional and chromatin landscape in MB memory tissue, examines how transcription in the MB is changed in response to memory forming environmental signals and highlights the consequences of losing an important chromatin regulator in memory neurons during acute memory processes. The high-quality datasets that have been generated in this thesis are a valuable resource to the scientific community and can be used to inform many new lines of inquiry. For example, there are over 130 epigenetic regulators currently implicated in neurodevelopmental disorders, of which Bap60/SMARCD1 is one (Ciptasari & van Bokhoven, 2020; Nixon et al., 2019). Many of these epigenetic regulators have only been studied in the context of development, but evidence is mounting that they also play a critical role in responding to environmental challenges (Merkling et al., 2015; Riahi et al., 2021). The training induced transcriptome that we generated fundamentally represents the brain adapting to changes in the environment. By contrasting the genes regulated during memory with the transcriptional programs that are misregulated after

impacting the function other disease-associated epigenetic regulators, target genes and pathways can be identified that provide mechanistic insight and develop new translational strategies for the associated diseases. This insight will be critical to help accelerate the move from model organism research to urgently needed patient treatment.

## Model of courtship memory TF hierarchy



**Figure 5.22 Developing a model of courtship memory transcription factor hierarchy**

Courtship memory signalling pathways converge to activate the constitutively expressed transcription factor CREB (*Crebb*). CREB goes on to activate the courtship IEGs Hr38 and sr. Together, CREB, Hr38 and sr go on to regulate an early wave of transcription in memory-specific and non-specific tissue, which play important roles in lipid metabolism, synaptic signaling and the oxidative stress response. The transcription factor sr specifically plays a role in activating the MB-specific memory transcriptional response, which includes important known memory signalling pathway components and tricarboxylic acid cycle metabolism, that is induced as courtship training progresses. The SWI/SNF complex is required for the inducibility of the MB-specific memory transcriptional response that is observed at the end of training.

## References

- Abascal, F., Acosta, R., Addleman, N. J., Adrian, J., Afzal, V., Aken, B., Ai, R., Akiyama, J. A., Jammal, O. Al, Amrhein, H., Anderson, S. M., Andrews, G. R., Antoshechkin, I., Ardlie, K. G., Armstrong, J., Astley, M., Banerjee, B., Barkal, A. A., Barnes, I. H. A., ... Myers, R. M. (2020). Perspectives on ENCODE. *Nature*, 583(7818), 693–698. <https://doi.org/10.1038/s41586-020-2449-8>
- Ackerman, S. L., & Siegel, R. W. (1986). Chemically Reinforced Conditioned Courtship in *Drosophila*: Responses of Wild-Type and the dunce, amnesiac and don giovanni Mutants. *Journal of Neurogenetics*, 3(2), 111–123. <https://doi.org/10.3109/01677068609106898>
- Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez-Banet, J., Billis, K., Garcia-Giron, C., Hourlier, T., Howe, K. L., Kahari, A. K., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, Y. A., ... Searle, S. M. J. (2016). The Ensembl Gene Annotation System. *Database : The Journal of Biological Databases and Curation*, 2016, baw093. <https://doi.org/10.1093/database/baw093>
- Al-Anzi, B., & Wyman, R. J. (2009). The *Drosophila* immunoglobulin gene turtle encodes guidance molecules involved in axon pathfinding. *Neural Development*, 4(1), 31. <https://doi.org/10.1186/1749-8104-4-31>
- Anders, S., Huber, W., Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., Mortazavi, A., Williams, B., McCue, K., Schaeffer, L., Wold, B., Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., ... Salzberg, S. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10), R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Aso, Y., Hattori, D., Yu, Y., Johnston, R. M., Iyer, N. A., Ngo, T. T. B., Dionne, H., Abbott, L. F., Axel, R., Tanimoto, H., & Rubin, G. M. (2014). The neuronal architecture of the mushroom body provides a logic for associative learning. *ELife*, 3, e04577. <https://doi.org/10.7554/eLife.04577>
- Barajas-Azpeleta, R., Wu, J., Gill, J., Welte, R., Seidel, C., McKinney, S., Dissel, S., & Si, K. (2018). Antimicrobial peptides modulate long-term memory. [https://www.lightningmaps.org/blitzortung/america/index.php?Bo\\_showmap=usaPLoS](https://www.lightningmaps.org/blitzortung/america/index.php?Bo_showmap=usaPLoS) *Genetics*, 14(10), 1–26. <https://doi.org/10.1371/journal.pgen.1007440>
- Biggar, S. R. (1999). Continuous and widespread roles for the Swi-Snf complex in transcription. *The EMBO Journal*, 18(8), 2254–2264. <https://doi.org/10.1093/emboj/18.8.2254>
- Blum, A. L., Li, W., Cressy, M., & Dubnau, J. (2009). Short- and long-term memory in *Drosophila* require cAMP signaling in distinct neuron types. *Current Biology : CB*, 19(16), 1341–1350. <https://doi.org/10.1016/j.cub.2009.07.016>
- Bolduc, F. V., & Tully, T. (2009). Fruit flies and intellectual disability. *Fly*, 3(1), 91–104. <https://doi.org/10.4161/fly.3.1.7812>

- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Brand, A. H., & Perrimon, N. (1993). *Targeted gene expression as a means of altering cell fates and generating dominant phenotypes*. 415, 401–415.
- Broughton, S. J., Tully, T., & Greenspan, R. J. (2003). Conditioning deficits of CaM-kinase transgenic *Drosophila melanogaster* in a new excitatory courtship assay. *Journal of Neurogenetics*, 17(1), 91–102. <http://www.ncbi.nlm.nih.gov/pubmed/14504030>
- Brunelli, M., Castellucci, V., & Kandel, E. R. (1976). Synaptic facilitation and behavioral sensitization in *Aplysia*: possible role of serotonin and cyclic AMP. *Science (New York, N.Y.)*, 194(4270), 1178–1181. <https://doi.org/10.1126/science.186870>
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology*, 109(1), 1–10. <https://doi.org/10.1002/0471142727.mb2129s109>
- Chen, X., Rahman, R., Guo, F., & Rosbash, M. (2016). Genome-wide identification of neuronal activity-regulated genes in *Drosophila*. *ELife*, 5, 1–21. <https://doi.org/10.7554/eLife.19942>
- Chubak, M. C., Nixon, K. C. J., Stone, M. H., Raun, N., Rice, S. L., Sarikahya, M., Jones, S. G., Lyons, T. A., Jakub, T. E., Mainland, R. L. M., Knip, M. J., Edwards, T. N., & Kramer, J. M. (2019). Individual components of the SWI/SNF chromatin remodelling complex have distinct roles in memory neurons of the *Drosophila* mushroom body. *Disease Models & Mechanisms*, 12(3). <https://doi.org/10.1242/dmm.037325>
- Ciptasari, U., & van Bokhoven, H. (2020). The phenomenal epigenome in neurodevelopmental disorders. *Human Molecular Genetics*, 29(R1), R42–R50. <https://doi.org/10.1093/hmg/ddaa175>
- Clapier, C. R., & Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, 78(1), 273–304. <https://doi.org/10.1146/annurev.biochem.77.062706.153223>
- Clowney, E. J., Iguchi, S., Bussell, J. J., Scheer, E., & Ruta, V. (2015). Multimodal Chemosensory Circuits Controlling Male Courtship in *Drosophila*. *Neuron*, 87(5), 1036–1049. <https://doi.org/10.1016/j.neuron.2015.07.025>
- Coll-Tane, M., Krebbers, A., Castells-Nobau, A., Zweier, C., & Schenck, A. (2019). Intellectual disability and autism spectrum disorders “on the fly”: Insights from *Drosophila*. *DMM Disease Models and Mechanisms*, 12(5). <https://doi.org/10.1242/dmm.039180>
- Connolly, J. B., Roberts, I. J. H., Armstrong, J. D., Kaiser, K., Forte, M., Tully, T., & O’Kane, C. J. (1996). Associative Learning Disrupted by Impaired G s Signaling in *Drosophila* Mushroom Bodies. *Science*, 274(5295), 2104–2107. <https://doi.org/10.1126/science.274.5295.2104>
- Dag, U., Lei, Z., Le, J. Q., Wong, A., Bushey, D., & Keleman, K. (2019). Neuronal reactivation during post-learning sleep consolidates long-term memory in *Drosophila*. *ELife*, 8, 1–23. <https://doi.org/10.7554/eLife.42786>

- Dal Molin, A., & Di Camillo, B. (2019). How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives. *Briefings in Bioinformatics*, 20(4), 1384–1394. <https://doi.org/10.1093/bib/bby007>
- Davis, R. L. (2011). Traces of Drosophila Memory. *Neuron*, 70(1), 8–19. <https://doi.org/10.1016/j.neuron.2011.03.012>
- de Belle, J. S., & Heisenberg, M. (1994). Associative odor learning in Drosophila abolished by chemical ablation of mushroom bodies. In *Science* (Vol. 263, Issue 5147, pp. 692–695). <https://doi.org/10.1126/science.8303280>
- de Tredern, E., Rabah, Y., Pasquer, L., Minatchy, J., Plaçais, P. Y., & Preat, T. (2021). Glial glucose fuels the neuronal pentose phosphate pathway for long-term memory. *Cell Reports*, 36(8). <https://doi.org/10.1016/j.celrep.2021.109620>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Domjan, M. (2005). Pavlovian Conditioning: A Functional Perspective. *Annual Review of Psychology*, 56(1), 179–206. <https://doi.org/10.1146/annurev.psych.55.090902.141409>
- Drain, P., Folkers, E., & Quinn, W. C. (1991). CAMP-Dependent Protein of learning in Transgenic Kinase and the Disruption Flies. *Cell Press*, 6, 71–82. [https://doi.org/https://doi.org/10.1016/0896-6273\(91\)90123-H](https://doi.org/https://doi.org/10.1016/0896-6273(91)90123-H)
- Dubnau, J., Chiang, A. S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U., Broger, C., & Tully, T. (2003). The staufen/pumilio pathway is involved in drosophila long-term memory. *Current Biology*, 13(4), 286–296. [https://doi.org/10.1016/S0960-9822\(03\)00064-2](https://doi.org/10.1016/S0960-9822(03)00064-2)
- Dudai, Y., Jan, Y. N., Byers, D., Quinn, W. G., & Benzer, S. (1976). dunce, a mutant of Drosophila deficient in learning. *Proceedings of the National Academy of Sciences of the United States of America*, 73(5), 1684–1688. <https://doi.org/10.1073/pnas.73.5.1684>
- Ejima, A., Smith, B. P. C., Lucas, C., Levine, J. D., & Griffith, L. C. (2005). Sequential Learning of Pheromonal Cues Modulates Memory Consolidation in Trainer-Specific Associative Courtship Conditioning. *Current Biology*, 15(3), 194–206. <https://doi.org/10.1016/j.cub.2005.01.035>
- Ejima, A., Smith, B. P. C., Lucas, C., van der Goes van Naters, W., Miller, C. J., Carlson, J. R., Levine, J. D., & Griffith, L. C. (2007). Generalization of Courtship Learning in Drosophila Is Mediated by cis-Vaccenyl Acetate. *Current Biology*, 17(7), 599–605. <https://doi.org/10.1016/j.cub.2007.01.053>
- ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. <https://doi.org/10.1038/nature11247>
- Fujita, N., Nagata, Y., Nishiuchi, T., Sato, M., Iwami, M., & Kiya, T. (2013). Visualization of neural activity in insect brains using a conserved immediate early gene, Hr38. *Current Biology*, 23(20), 2063–2070. <https://doi.org/10.1016/j.cub.2013.08.051>

- Gailey, D. A., Jackson, F. R., & Siegel, R. W. (1984). Conditioning Mutations in DROSOPHILA MELANOGASTER Affect an Experience-Dependent Behavioral Modification in Courting Males. *Genetics*, *106*(4), 613–61323.
- Gibson, G. (2022). Perspectives on rigor and reproducibility in single cell genomics. *PLOS Genetics*, *18*(5), e1010210. <https://doi.org/10.1371/journal.pgen.1010210>
- Gil-Marti, B., Barredo, C. G., Pina-Flores, S., Trejo, J. L., Turiegano, E., & Martin, F. A. (2022). The elusive transcriptional memory trace. *Oxford Open Neuroscience*, *1*(June), 1–8. <https://doi.org/10.1093/oons/kvac008>
- Giorgi, C., & Marinelli, S. (2021). Roles and Transcriptional Responses of Inhibitory Neurons in Learning and Memory. *Frontiers in Molecular Neuroscience*, *14*. <https://doi.org/10.3389/fnmol.2021.689952>
- Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. *Cell*, *128*(4), 635–638. <https://doi.org/10.1016/j.cell.2007.02.006>
- Goodwin, S. F., Del Vecchio, M., Velinzon, K., Hogel, C., Russell, S. R. H., Tully, T., & Kaiser, K. (1997). Defective Learning in Mutants of the Drosophila Gene for a Regulatory Subunit of cAMP-Dependent Protein Kinase. *The Journal of Neuroscience*, *17*(22), 8817–8827. <https://doi.org/10.1523/JNEUROSCI.17-22-08817.1997>
- Grosjean, Y., Guenin, L., Bardet, H. M., & Ferveur, J. F. (2007). Prospero mutants induce precocious sexual behavior in Drosophila males. *Behavior Genetics*, *37*(4), 575–584. <https://doi.org/10.1007/s10519-007-9152-5>
- Grosjean, Y., Savy, M., Soichot, J., Everaerts, C., Cézilly, F., & Ferveur, J.-F. (2004). Mild mutations in the pan neural gene prospero affect male-specific behaviour in Drosophila melanogaster. *Behavioural Processes*, *65*(1), 7–13. [https://doi.org/10.1016/S0376-6357\(03\)00148-7](https://doi.org/10.1016/S0376-6357(03)00148-7)
- Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review*, *77*(5), 419–450. <https://doi.org/10.1037/h0029810>
- Halder, R., Hennion, M., Vidal, R. O., Shomroni, O., Rahman, R.-U., Rajput, A., Centeno, T. P., van Bebber, F., Capece, V., Vizcaino, J. C. G., Schuetz, A.-L., Burkhardt, S., Benito, E., Sala, M. N., Javan, S. B., Haass, C., Schmid, B., Fischer, A., & Bonn, S. (2016). DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. *Nature Neuroscience*, *19*(1), 102–110. <https://doi.org/10.1038/nn.4194>
- Han, P.-L., Levin, L. R., Reed, R. R., & Davis, R. L. (1992). Preferential expression of the drosophila rutabaga gene in mushroom bodies, neural centers for learning in insects. *Neuron*, *9*(4), 619–627. [https://doi.org/10.1016/0896-6273\(92\)90026-A](https://doi.org/10.1016/0896-6273(92)90026-A)
- Hargreaves, D. C., & Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Research*, *21*(3), 396–420. <https://doi.org/10.1038/cr.2011.32>
- Hawkins, R. D., Kandel, E. R., & Bailey, C. H. (2006). Molecular mechanisms of memory storage in Aplysia. *The Biological Bulletin*, *210*(3), 174–191. <https://doi.org/210/3/174> [pii]

- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, 38(4), 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>
- Heisenberg, M., Borst, A., Wagner, S., & Byers, D. (1985). Drosophila Mushroom Body Mutants are Deficient in Olfactory Learning. *Journal of Neurogenetics*, 2(1), 1–30. <https://doi.org/10.3109/01677068509100140>
- Henikoff, S., Henikoff, J. G., Kaya-Okur, H. S., & Ahmad, K. (2020). Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *ELife*, 9. <https://doi.org/10.7554/eLife.63274>
- Henry, G. L., Davis, F. P., Picard, S., & Eddy, S. R. (2012). Cell type-specific genomics of Drosophila neurons. *Nucleic Acids Research*, 40(19), 9691–9704. <https://doi.org/10.1093/nar/gks671>
- Hirano, Y., Ihara, K., Masuda, T., Yamamoto, T., Iwata, I., Takahashi, A., Awata, H., Nakamura, N., Takakura, M., Suzuki, Y., Horiuchi, J., Okuno, H., & Saitoe, M. (2016). Shifting transcriptional machinery is required for long-term memory maintenance and modification in Drosophila mushroom bodies. *Nature Communications*, 7, 13471. <https://doi.org/10.1038/ncomms13471>
- Hrvatin, S., Hochbaum, D. R., Nagy, M. A., Cicconet, M., Robertson, K., Cheadle, L., Zilionis, R., Ratner, A., Borges-Monroy, R., Klein, A. M., Sabatini, B. L., & Greenberg, M. E. (2018). Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nature Neuroscience*, 21(1), 120–129. <https://doi.org/10.1038/s41593-017-0029-5>
- Hulsen, T., de Vlieg, J., & Alkema, W. (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, 9, 488. <https://doi.org/10.1186/1471-2164-9-488>
- Ishimoto, H., Wang, Z., Rao, Y., Wu, C. F., & Kitamoto, T. (2013). A Novel Role for Ecdysone in Drosophila Conditioned Behavior: Linking GPCR-Mediated Non-canonical Steroid Action to cAMP Signaling in the Adult Brain. *PLoS Genetics*, 9(10). <https://doi.org/10.1371/journal.pgen.1003843>
- Iyer, E. P. R., Iyer, S. C., Sullivan, L., Wang, D., Meduri, R., Graybeal, L. L., & Cox, D. N. (2013). Functional Genomic Analyses of Two Morphologically Distinct Classes of Drosophila Sensory Neurons: Post-Mitotic Roles of Transcription Factors in Dendritic Patterning. *PLoS ONE*, 8(8), e72434. <https://doi.org/10.1371/journal.pone.0072434>
- Jakub, T., Quesnel, K., Keung, C., Bérubé, N. G., & Kramer, J. M. (2021). Epigenetics in intellectual disability. In *Epigenetics in Psychiatry* (pp. 489–517). Elsevier. <https://doi.org/10.1016/B978-0-12-823577-5.00030-1>
- Jauregui-Lozano, J., Bakhle, K., & Weake, V. M. (2021). In vivo tissue-specific chromatin profiling in Drosophila melanogaster using GFP-tagged nuclei. *Genetics*, 218(3). <https://doi.org/10.1093/genetics/iyab079>

- Jenett, A., Rubin, G. M., Ngo, T. T. B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J., Iyer, N., Fetter, D., Hausenfluck, J. H., Peng, H., Trautman, E. T., Svirskas, R. R., Myers, E. W., Iwinski, Z. R., Aso, Y., ... Zugates, C. T. (2012). A GAL4-Driver Line Resource for Drosophila Neurobiology. *Cell Reports*, 2(4), 991–1001. <https://doi.org/10.1016/j.celrep.2012.09.011>
- Joiner, M. A., & Griffith, L. C. (1997). CaM Kinase II and Visual Input Modulate Memory Formation in the Neuronal Circuit Controlling Courtship Conditioning. *The Journal of Neuroscience*, 17(23), 9384–9391. <https://doi.org/10.1523/JNEUROSCI.17-23-09384.1997>
- Jones, S. G., Nixon, K. C. J., Chubak, M. C., & Kramer, J. M. (2018). Mushroom Body Specific Transcriptome Analysis Reveals Dynamic Regulation of Learning and Memory Genes After Acquisition of Long-Term Courtship Memory in Drosophila. *Genes/Genomes/Genetics*, 8(November), 3433–3446. <https://doi.org/10.1534/g3.118.200560>
- Kaldun, J. C., & Sprecher, S. G. (2019). Initiated by CREB: Resolving Gene Regulatory Programs in Learning and Memory. *BioEssays*, 41(8), 1900045. <https://doi.org/10.1002/bies.201900045>
- Kane, N. S., Robichon, A., Dickinson, J. A., & Greenspan, R. J. (1997). Learning without Performance in PKC-Deficient Drosophila. *Neuron*, 18(2), 307–314. [https://doi.org/10.1016/S0896-6273\(00\)80270-6](https://doi.org/10.1016/S0896-6273(00)80270-6)
- Kassabov, S. R., Zhang, B., Persinger, J., & Bartholomew, B. (2003). SWI/SNF Unwraps, Slides, and Rewraps the Nucleosome. *Molecular Cell*, 11(2), 391–403. [https://doi.org/10.1016/S1097-2765\(03\)00039-X](https://doi.org/10.1016/S1097-2765(03)00039-X)
- Keegan, S. E., Haskins, J., Simmonds, A. J., & Hughes, S. C. (2023). A chromatin remodelling SWI/SNF subunit Snr1 regulates neural stem cell determination and differentiation. *Development*. <https://doi.org/10.1242/dev.201484>
- Keleman, K., Krüttner, S., Alenius, M., & Dickson, B. J. (2007). Function of the Drosophila CPEB protein Orb2 in long-term courtship memory. *Nature Neuroscience*, 10(12), 1587–1593. <https://doi.org/10.1038/nn1996>
- Keleman, K., Vrontou, E., Krüttner, S., Yu, J. Y., Kurtovic-Kozaric, A., & Dickson, B. J. (2012). Dopamine neurons modulate pheromone responses in Drosophila courtship learning. *Nature*, 489(7414), 145–149. <https://doi.org/10.1038/nature11345>
- Kirilly, D., Wong, J. J. L., Lim, E. K. H., Wang, Y., Zhang, H., Wang, C., Liao, Q., Wang, H., Liou, Y.-C., Wang, H., & Yu, F. (2011). Intrinsic Epigenetic Factors Cooperate with the Steroid Hormone Ecdysone to Govern Dendrite Pruning in Drosophila. *Neuron*, 72(1), 86–100. <https://doi.org/10.1016/j.neuron.2011.08.003>
- Koemans, T. S., Kleefstra, T., Chubak, M. C., Stone, M. H., Reijnders, M. R. F., de Munnik, S., Willemsen, M. H., Fenckova, M., Stumpel, C. T. R. M., Bok, L. A., Sifuentes Saenz, M., Byerly, K. A., Baughn, L. B., Stegmann, A. P. A., Pfundt, R., Zhou, H., van Bokhoven, H., Schenck, A., & Kramer, J. M. (2017). Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. *PLOS Genetics*, 13(10), e1006864. <https://doi.org/10.1371/journal.pgen.1006864>

- Koemans, T. S., Oppitz, C., Donders, R. A. T., van Bokhoven, H., Schenck, A., Keleman, K., & Kramer, J. M. (2017). *Drosophila* Courtship Conditioning As a Measure of Learning and Memory. *Journal of Visualized Experiments*, *124*, 1–11. <https://doi.org/10.3791/55808>
- Krashes, M. J., Keene, A. C., Leung, B., Armstrong, J. D., & Waddell, S. (2007). Sequential Use of Mushroom Body Neuron Subsets during *Drosophila* Odor Memory Processing. *Neuron*, *53*(1), 103–115. <https://doi.org/10.1016/j.neuron.2006.11.021>
- Krüttner, S., Traunmüller, L., Dag, U., Jandrasits, K., Stepien, B., Iyer, N., Fradkin, L. G., Noordermeer, J. N., Mensh, B. D., & Keleman, K. (2015). Synaptic Orb2A Bridges Memory Acquisition and Late Memory Consolidation in *Drosophila*. *Cell Reports*, *11*(12), 1953–1965. <https://doi.org/10.1016/j.celrep.2015.05.037>
- Kummeling, J., Stremmelaar, D. E., Raun, N., Reijnders, M. R. F., Willemsen, M. H., Ruitkamp-Versteeg, M., Schepens, M., Man, C. C. O., Gilissen, C., Cho, M. T., McWalter, K., Sinnema, M., Wheless, J. W., Simon, M. E. H., Genetti, C. A., Casey, A. M., Terhal, P. A., van der Smagt, J. J., van Gassen, K. L. I., ... Kleefstra, T. (2021). Characterization of SETD1A haploinsufficiency in humans and *Drosophila* defines a novel neurodevelopmental syndrome. *Molecular Psychiatry*, *26*(6), 2013–2024. <https://doi.org/10.1038/s41380-020-0725-5>
- Langmead, B., & Salzberg, S. L. (2012). *Fast gapped-read alignment with Bowtie 2*. *9*(4), 357–360. <https://doi.org/10.1038/nmeth.1923>
- Lau, H. L., Timbers, T. A., Mahmoud, R., & Rankin, C. H. (2013). Genetic dissection of memory for associative and non-associative learning in *Caenorhabditis elegans*. *Genes, Brain and Behavior*, *12*(2), 210–223. <https://doi.org/10.1111/j.1601-183X.2012.00863.x>
- Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Stahl, B. T., Wu, H., Aebersold, R., Graef, I. A., & Crabtree, G. R. (2007). An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron*, *55*(2), 201–215. <https://doi.org/10.1016/j.neuron.2007.06.019>
- Levin, L. R., Han, P. L., Hwang, P. M., Feinstein, P. G., Davis, R. L., & Reed, R. R. (1992). The *Drosophila* learning and memory gene *rutabaga* encodes a Ca<sup>2+</sup> calmodulin-responsive adenylyl cyclase. *Cell*, *68*(3), 479–489. [https://doi.org/10.1016/0092-8674\(92\)90185-F](https://doi.org/10.1016/0092-8674(92)90185-F)
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Data, G. P., & Sam, T. (2009). *The Sequence Alignment / Map format and SAMtools*. *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, K. Le, Zhang, L., Yang, X. M., Fang, Q., Yin, X. F., Wei, H. M., Zhou, T., Li, Y. Bin, Chen, X. L., Tang, F., Li, Y. H., Chang, J. F., Li, W., & Sun, F. (2018). Histone acetyltransferase CBP-related H3K23 acetylation contributes to courtship learning in *Drosophila*. *06 Biological Sciences 0604 Genetics. BMC Developmental Biology*, *18*(1), 1–14. <https://doi.org/10.1186/s12861-018-0179-z>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). Sequence analysis featureCounts : an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>

- Livingstone, M. S., Sziber, P. P., & Quinn, W. G. (1984). Loss of Calcium Calmodulin Responsiveness in Adenylate-Cyclase of Rutabaga, a Drosophila Learning Mutant. *Cell*, 37(1), 205–215.
- Lomvardas, S., & Thanos, D. (2001). Nucleosome Sliding via TBP DNA Binding In Vivo. *Cell*, 106(6), 685–696. [https://doi.org/10.1016/S0092-8674\(01\)00490-1](https://doi.org/10.1016/S0092-8674(01)00490-1)
- Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., & Manke, T. (2021). pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics*, 37(3), 422–423. <https://doi.org/10.1093/bioinformatics/btaa692>
- López, A. J., & Wood, M. A. (2015). Role of nucleosome remodeling in neurodevelopmental and intellectual disability disorders. *Frontiers in Behavioral Neuroscience*, 9. <https://doi.org/10.3389/fnbeh.2015.00100>
- Lorch, Y., Zhang, M., & Kornberg, R. D. (2001). RSC Unravels the Nucleosome. *Molecular Cell*, 7(1), 89–95. [https://doi.org/10.1016/S1097-2765\(01\)00157-5](https://doi.org/10.1016/S1097-2765(01)00157-5)
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/s13059-014-0550-8>
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), 251–260. <https://doi.org/10.1038/38444>
- Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., Baymuradov, U. K., Graham, K., Litton, C., Miyasato, S. R., Strattan, J. S., Jolanki, O., Lee, J.-W., Tanaka, F. Y., Adenekan, P., ... Cherry, J. M. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Research*, 48(D1), D882–D889. <https://doi.org/10.1093/nar/gkz1062>
- Machado Almeida, P., Lago Solis, B., Stickle, L., Feidler, A., & Nagoshi, E. (2021). Neurofibromin 1 in mushroom body neurons mediates circadian wake drive through activating cAMP–PKA signaling. *Nature Communications*, 12(1), 1–17. <https://doi.org/10.1038/s41467-021-26031-2>
- Mardinly, A. R., Spiegel, I., Patrizi, A., Centofante, E., Bazinet, J. E., Tzeng, C. P., Mandel-Brehm, C., Harmin, D. A., Adesnik, H., Fagiolini, M., & Greenberg, M. E. (2016). Sensory experience regulates cortical inhibition by inducing IGF1 in VIP neurons. *Nature*, 531(7594), 371–375. <https://doi.org/10.1038/nature17187>
- Marshall, O. J., Southall, T. D., Cheetham, S. W., & Brand, A. H. (2016). Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nature Protocols*, 11(9), 1586–1598. <https://doi.org/10.1038/nprot.2016.084>
- Mashtalir, N., D’Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., Zullo, H. J., McKenzie, Z. M., Kubiak, R. L., St. Pierre, R., Valencia, A. M., Poynter, S. J., Cassel, S. H., Ranish, J. A., & Kadoch, C. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell*, 1–17. <https://doi.org/10.1016/j.cell.2018.09.032>

- McBride, S. M. ., Giuliani, G., Choi, C., Krause, P., Correale, D., Watson, K., Baker, G., & Siwicki, K. K. (1999). Mushroom Body Ablation Impairs Short-Term Memory and Long-Term Memory of Courtship Conditioning in *Drosophila melanogaster*. *Neuron*, *24*(4), 967–977. [https://doi.org/10.1016/S0896-6273\(00\)81043-0](https://doi.org/10.1016/S0896-6273(00)81043-0)
- Mcguire, S. E., Le, P. T., Osborn, A. J., Matsumoto, K., & Davis, R. L. (2003). Spatiotemporal Rescue of Memory Dysfunction in *Drosophila*. *Science*, *302*(December), 1765–1769. <https://doi.org/10.1126/science.1089035>
- Mehren, J. E., & Griffith, L. C. (2006). Cholinergic neurons mediate CaMKII-dependent enhancement of courtship suppression. *Learning & Memory*, *13*(6), 686–689. <https://doi.org/10.1101/lm.317806>
- Merkling, S. H., Bronkhorst, A. W., Kramer, J. M., Overheul, G. J., Schenck, A., & Van Rij, R. P. (2015). The Epigenetic Regulator G9a Mediates Tolerance to RNA Virus Infection in *Drosophila*. *PLoS Pathogens*, *11*(4), 1–25. <https://doi.org/10.1371/journal.ppat.1004692>
- Miller, C. A., Gavin, C. F., White, J. A., Parrish, R. R., Honasoge, A., Yancey, C. R., Rivera, I. M., Rubio, M. D., Rumbaugh, G., & Sweatt, J. D. (2010). Cortical DNA methylation maintains remote memory. *Nature Neuroscience*, *13*(6), 664–666. <https://doi.org/10.1038/nn.2560>
- Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive Neuroscience and the Study of Memory. *Neuron*, *20*(3), 445–468. [https://doi.org/10.1016/S0896-6273\(00\)80987-3](https://doi.org/10.1016/S0896-6273(00)80987-3)
- Montague, S. A., & Baker, B. S. (2016). Memory elicited by courtship conditioning requires mushroom body neuronal subsets similar to those utilized in appetitive memory. *PLoS ONE*, *11*(10), 1–24. <https://doi.org/10.1371/journal.pone.0164516>
- Montarolo, P. G., Goelet, P., Castellucci, V. F., Morgan, J., Kandel, E. R., & Schacher, S. (1986). A critical period for macromolecular synthesis in long-term heterosynaptic facilitation in *Aplysia*. *Science (New York, N.Y.)*, *234*(4781), 1249–1254. <https://doi.org/10.1126/science.3775383>
- Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G., & Goodman, R. H. (1986). Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(18), 6682–6686. <https://doi.org/10.1073/pnas.83.18.6682>
- Morgan, T. H. (1910). Sex Limited Inheritance in *Drosophila*. *Science*, *32*(812), 120–122. <https://doi.org/10.1126/science.32.812.120>
- Narayanan, R., Pirouz, M., Kerimoglu, C., Pham, L., Wagener, R. J., Kiszka, K. A., Rosenbusch, J., Seong, R. H., Kessel, M., Fischer, A., Stoykova, A., Staiger, J. F., & Tuoc, T. (2015). Loss of BAF (mSWI/SNF) Complexes Causes Global Transcriptional and Chromatin State Changes in Forebrain Development. *Cell Reports*, *13*(9), 1842–1854. <https://doi.org/10.1016/j.celrep.2015.10.046>
- Narayanan, R., & Tuoc, T. C. (2014). Roles of chromatin remodeling BAF complex in neural differentiation and reprogramming. *Cell and Tissue Research*, *356*(3), 575–584. <https://doi.org/10.1007/s00441-013-1791-7>

- Narkaj, K., Stefanelli, G., Wahdan, M., Azam, A. B., Ramzan, F., Steininger, C. F. D., Walters, B. J., & Zovkic, I. B. (2018). Blocking H2A.Z Incorporation via Tip60 Inhibition Promotes Systems Consolidation of Fear Memory in Mice. *Eneuro*, *5*(5), ENEURO.0378-18.2018. <https://doi.org/10.1523/ENEURO.0378-18.2018>
- Neugeborn, L., & Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics*, *108*(4), 845–858. <https://doi.org/10.1093/genetics/108.4.845>
- Nixon, K. C. J. (2020). Mushroom body-specific gene regulation by the SWI/SNF chromatin remodeling complex. In *Electronic Thesis and Dissertation Repository*. <https://ir.lib.uwo.ca/etd/6818>
- Nixon, K. C. J., Rousseau, J., Stone, M. H., Sarikahya, M., Ehresmann, S., Mizuno, S., Matsumoto, N., Miyake, N., Study, D. D. D., Baralle, D., Mckee, S., Izumi, K., Ritter, A. L., Heide, S., He, D., Depienne, C., Titheradge, H., & Kramer, J. M. (2019). A Syndromic Neurodevelopmental Disorder Caused by Mutations in SMARCD1 , a Core SWI / SNF Subunit Needed for Context-Dependent Neuronal Gene Regulation in Flies. *American Journal of Human Genetics*, *104*, 596–610. <https://doi.org/10.1016/j.ajhg.2019.02.001>
- O'dell, K. M. C., Jamieson, D., Goodwin, S. F., & Kaiser, K. (1999). Abnormal Courtship Conditioning in Males Mutant for the RI Regulatory Subunit of Drosophila Protein Kinase A. *Journal of Neurogenetics*, *13*(1–2), 105–118. <https://doi.org/10.3109/01677069909083469>
- Olave, I., Wang, W., Xue, Y., Kuo, A., & Crabtree, G. R. (2002). Identification of a polymorphic, neuron-specific chromatin remodeling complex. *Genes & Development*, *16*(19), 2509–2517. <https://doi.org/10.1101/gad.992102>
- Parrish, J. Z., Kim, M. D., Jan, L. Y., & Jan, Y. N. (2006). Genome-wide analyses identify transcription factors required for proper morphogenesis of Drosophila sensory neuron dendrites. *Genes & Development*, *20*(7), 820–835. <https://doi.org/10.1101/gad.1391006>
- Pavlou, H. J., & Goodwin, S. F. (2013). Courtship behavior in *Drosophila melanogaster*: Towards a “courtship connectome.” *Current Opinion in Neurobiology*, *23*(1), 76–83. <https://doi.org/10.1016/j.conb.2012.09.002>
- Perkins, L. A., Holderbaum, L., Tao, R., Hu, Y., Sopko, R., McCall, K., Yang-Zhou, D., Flockhart, I., Binari, R., Shim, H.-S., Miller, A., Housden, A., Foos, M., Randkely, S., Kelley, C., Namgyal, P., Villalta, C., Liu, L.-P., Jiang, X., ... Perrimon, N. (2015). The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. *Genetics*, *201*(3), 843–852. <https://doi.org/10.1534/genetics.115.180208>
- Pinsker, H., Kupfermann, I., Castellucci, V., & Kandel, E. (1970). Habituation and Dishabituation of the Gill-Withdrawal Reflex in *Aplysia*. *Science*, *167*(3926), 1740–1742. <https://doi.org/10.1126/science.167.3926.1740>
- Plaçais, P. Y., De Tredern, É., Scheunemann, L., Trannoy, S., Goguel, V., Han, K. A., Isabel, G., & Preat, T. (2017). Upregulated energy metabolism in the *Drosophila* mushroom body is the trigger for long-term memory. *Nature Communications*, *8*(15510). <https://doi.org/10.1038/ncomms15510>

- Plaçais, P. Y., & Preat, T. (2013). To favor survival under food shortage, the brain disables costly memory. *Science*, 339(6118), 440–442. <https://doi.org/10.1126/science.1226018>
- Priam, P., Krasteva, V., Rousseau, P., D'Angelo, G., Gaboury, L., Sauvageau, G., & Lessard, J. A. (2017). SMARCD2 subunit of SWI/SNF chromatin-remodeling complexes mediates granulopoiesis through a CEBP $\epsilon$  dependent mechanism. *Nature Genetics*, 49(5), 753–764. <https://doi.org/10.1038/ng.3812>
- Quinn, W. G., Harris, W. A., & Benzer, S. (1974). Conditioned behavior in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 71(3), 708–712. <https://doi.org/VL-71>
- R Core Team. (2016). *R: A language and environment for statistical computing*. R foundation for statistical computing. <https://www.r-project.org/>
- Ram, F., Ryan, D. P., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Friederike, D., & Manke, T. (2016). *deepTools2: a next generation web server for deep-sequencing data analysis*. 44(April), 160–165. <https://doi.org/10.1093/nar/gkw257>
- Raun, N. (2019). The role of H3K4 methyltransferase in *Drosophila* memory. In *Electronic Thesis and Dissertation Repository*. <https://ir.lib.uwo.ca/etd/5993>
- Raun, N., Jones, S., & Kramer, J. M. (2021). Conditioned courtship suppression in *Drosophila melanogaster*. *Journal of Neurogenetics*, 0(0), 1–27. <https://doi.org/10.1080/01677063.2021.1873323>
- Reich, A., Sapir, A., & Shilo, B. (1999). Sprouty is a general inhibitor of receptor tyrosine kinase signaling. *Development (Cambridge, England)*, 126(18), 4139–4147. <https://doi.org/10.1242/dev.126.18.4139>
- Riahi, H., Fenckova, M., Goruk, K. J., Schenck, A., & Kramer, J. M. (2021). The epigenetic regulator G9a attenuates stress-induced resistance and metabolic transcriptional programs across different stressors and species. *BMC Biology*, 19(1), 1–15. <https://doi.org/10.1186/s12915-021-01025-0>
- Roote, J., & Prokop, A. (2013). How to Design a Genetic Mating Scheme: A Basic Training Package for *Drosophila* Genetics. *G3 Genes|Genomes|Genetics*, 3(2), 353–358. <https://doi.org/10.1534/g3.112.004820>
- Sakai, T., Tamura, T., Kitamoto, T., & Kidokoro, Y. (2004). A clock gene, period, plays a key role in long-term memory formation in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(45), 16058–16063. <https://doi.org/10.1073/pnas.0401472101>
- Schmidt, F., Gasparoni, N., Gasparoni, G., Gianmoena, K., Cadenas, C., Polansky, J. K., Ebert, P., Nordström, K., Barann, M., Sinha, A., Fröhler, S., Xiong, J., Dehghani Amirabad, A., Behjati Ardakani, F., Hutter, B., Zipprich, G., Felder, B., Eils, J., Brors, B., ... Schulz, M. H. (2017). Combining transcription factor binding affinities with open-chromatin data for accurate gene expression prediction. *Nucleic Acids Research*, 45(1), 54–66. <https://doi.org/10.1093/nar/gkw1061>

- Schones, D. E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., & Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, *132*(5), 887–898. <https://doi.org/10.1016/j.cell.2008.02.022>
- Schwaerzel, M., Monastirioti, M., Scholz, H., Friggi-Grelin, F., Birman, S., & Heisenberg, M. (2003). Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in *Drosophila*. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(33), 10495–10502. <http://www.ncbi.nlm.nih.gov/pubmed/14627633>
- Sgammeglia, N., & Sprecher, S. G. (2022). Interplay between metabolic energy regulation and memory pathways in *Drosophila*. *Trends in Neurosciences*, *45*(7), 539–549. <https://doi.org/10.1016/j.tins.2022.04.007>
- Sheng, M., Thompson, M. A., & Greenberg, M. E. (1991). CREB: a Ca<sup>2+</sup>-Regulated Transcription Factor Phosphorylated by Calmodulin-Dependent Kinases. *Science*, *252*(5011), 1427–1430. <https://doi.org/10.1126/science.1646483>
- Siegel, R. W., & Hall, J. C. (1979). Conditioned responses in courtship behavior of normal and mutant *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(7), 3430–3434. <https://doi.org/10.1073/pnas.76.7.3430>
- Silva, B., Mantha, O. L., Schor, J., Pascual, A., Plaçais, P. Y., Pavlowsky, A., & Preat, T. (2022). Glia fuel neurons with locally synthesized ketone bodies to sustain memory under starvation. *Nature Metabolism*, *4*(2), 213–224. <https://doi.org/10.1038/s42255-022-00528-6>
- Smolik, S. M., Rose, R. E., & Goodman, R. H. (1992). A cyclic AMP-responsive element-binding transcriptional activator in *Drosophila melanogaster*, dCREB-A, is a member of the leucine zipper family. *Molecular and Cellular Biology*, *12*(9), 4123–4131. <https://doi.org/10.1128/MCB.12.9.4123>. Updated
- Son, E. Y., & Crabtree, G. R. (2014). The role of BAF (mSWI/SNF) complexes in mammalian neural development. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics*, *166*(3), 333–349. <https://doi.org/10.1002/ajmg.c.31416>
- Southall, T. D., Gold, K. S., Egger, B., Davidson, C. M., Caygill, E. E., Marshall, O. J., & Brand, A. H. (2013). Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: Assaying RNA pol II occupancy in neural stem cells. *Developmental Cell*, *26*(1), 101–112. <https://doi.org/10.1016/j.devcel.2013.05.020>
- Spieth, H. T. (1974). Courtship Behavior in *Drosophila*. *Annual Review of Entomology*, *19*(1), 385–405. <https://doi.org/10.1146/annurev.en.19.010174.002125>
- Squire, L. R. (2009). Memory and Brain Systems: 1969–2009. *The Journal of Neuroscience*, *29*(41), 12711–12716. <https://doi.org/10.1523/JNEUROSCI.3575-09.2009>
- Stark, R., & Brown, G. (2011). *DiffBind: differential binding analysis of ChIP-Seq peak data*. <http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf>
- Steensel, B. van, & Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nature Biotechnology*, *18*(4), 424–428. <https://doi.org/10.1038/74487>

- Stern, M., Jensen, R., & Herskowitz, I. (1984). Five SWI genes are required for expression of the HO gene in yeast. *Journal of Molecular Biology*, 178(4), 853–868. [https://doi.org/10.1016/0022-2836\(84\)90315-2](https://doi.org/10.1016/0022-2836(84)90315-2)
- Sweatt, J. D. (2010). *Mechanisms of memory* (2nd editio). Elsevier Academic Press.
- Takayanagi-kiya, S., & Kiya, T. (2019). *Activity-dependent visualization and control of neural circuits for courtship behavior in the fly Drosophila melanogaster*. 116(12), 5715–5720. <https://doi.org/10.1073/pnas.1814628116>
- Tea, J. S., & Luo, L. (2011). The chromatin remodeling factor Bap55 functions through the TIP60 complex to regulate olfactory projection neuron dendrite targeting. *Neural Development*, 6(1), 5. <https://doi.org/10.1186/1749-8104-6-5>
- Tian, Y., Zhang, Z. C., & Han, J. (2017). Drosophila Studies on Autism Spectrum Disorders. *Neuroscience Bulletin*, 33(6), 737–746. <https://doi.org/10.1007/s12264-017-0166-6>
- Trannoy, S., Redt-Clouet, C., Dura, J. M., & Preat, T. (2011). Parallel processing of appetitive short- and long-term memories in Drosophila. *Current Biology*, 21(19), 1647–1653. <https://doi.org/10.1016/j.cub.2011.08.032>
- Tully, T., Preat, T., Boynton, S. C., & Del Vecchio, M. (1994). Genetic dissection of consolidated memory in Drosophila. *Cell*, 79(1), 35–47. [https://doi.org/10.1016/0092-8674\(94\)90398-0](https://doi.org/10.1016/0092-8674(94)90398-0)
- Tully, Tim, Bourtchouladze, R., Scott, R., & Tallman, J. (2003). Targeting the CREB pathway for memory enhancers. *Nature Reviews. Drug Discovery*, 2(4), 267–277. <https://doi.org/10.1038/nrd1061>
- Tuoc, T., Dere, E., Radyushkin, K., Pham, L., Nguyen, H., Tonchev, A. B., Sun, G., Ronnenberg, A., Shi, Y., Staiger, J. F., Ehrenreich, H., & Stoykova, A. (2017). Ablation of BAF170 in Developing and Postnatal Dentate Gyrus Affects Neural Stem Cell Proliferation, Differentiation, and Learning. *Molecular Neurobiology*, 54(6), 4618–4635. <https://doi.org/10.1007/s12035-016-9948-5>
- Vierbuchen, T., Ling, E., Cowley, C. J., Couch, C. H., Wang, X., Harmin, D. A., Roberts, C. W. M., & Greenberg, M. E. (2017). AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Molecular Cell*, 68(6), 1134–1146.e6. <https://doi.org/10.1016/j.molcel.2017.11.026>
- Vogel-Ciernia, A., Matheos, D. P., Barrett, R. M., Kramár, E. A., Azzawi, S., Chen, Y., Magnan, C. N., Zeller, M., Sylvain, A., Haettig, J., Jia, Y., Tran, A., Dang, R., Post, R. J., Chabrier, M., Babayan, A. H., Wu, J. I., Crabtree, G. R., Baldi, P., ... Wood, M. A. (2013a). The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nature Neuroscience*, 16(5), 552–561. <https://doi.org/10.1038/nn.3359>
- Vogel-Ciernia, A., Matheos, D. P., Barrett, R. M., Kramár, E. A., Azzawi, S., Chen, Y., Magnan, C. N., Zeller, M., Sylvain, A., Haettig, J., Jia, Y., Tran, A., Dang, R., Post, R. J., Chabrier, M., Babayan, A. H., Wu, J. I., Crabtree, G. R., Baldi, P., ... Wood, M. A. (2013b). The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nature Neuroscience*, 16(5), 552–561. <https://doi.org/10.1038/nn.3359>

- Wang, Q., Li, M., Wu, T., Zhan, L., Li, L., Chen, M., Xie, W., Xie, Z., Hu, E., Xu, S., & Yu, G. (2022). Exploring Epigenomic Datasets by ChIPseeker. *Current Protocols*, 2(10). <https://doi.org/10.1002/cpz1.585>
- Winbush, A., Reed, D., Chang, P. L., Nuzhdin, S. V., Lyons, L. C., & Arbeitman, M. N. (2012). Identification of gene expression changes associated with long-term memory of courtship rejection in *Drosophila* males. *G3 (Bethesda, Md.)*, 2(11), 1437–1445. <https://doi.org/10.1534/g3.112.004119>
- Wu, J. I., Lessard, J., Olave, I. A., Qiu, Z., Ghosh, A., Graef, I. A., & Crabtree, G. R. (2007). Regulation of Dendritic Development by Neuron-Specific Chromatin Remodeling Complexes. *Neuron*, 56(1), 94–108. <https://doi.org/10.1016/j.neuron.2007.08.021>
- Wu, Y. E., Pan, L., Zuo, Y., Li, X., & Hong, W. (2017). Detecting Activated Cell Populations Using Single-Cell RNA-Seq. *Neuron*, 96(2), 313–329.e6. <https://doi.org/10.1016/j.neuron.2017.09.026>
- Yan, F., Powell, D. R., Curtis, D. J., & Wong, N. C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biology*, 21(1), 22. <https://doi.org/10.1186/s13059-020-1929-3>
- Yap, E., & Greenberg, M. E. (2018). Review Activity-Regulated Transcription : Bridging the Gap between Neural Activity and Behavior. *Neuron*, 100(2), 330–348. <https://doi.org/10.1016/j.neuron.2018.10.013>
- Yheskel, M., Sidoli, S., & Secombe, J. (2023). Proximity labeling reveals a new in vivo network of interactors for the histone demethylase KDM5. *Epigenetics & Chromatin*, 16(1), 8. <https://doi.org/10.1186/s13072-023-00481-y>
- Yin, J. C. P., Wallach, J. S., Del Vecchio, M., Wilder, E. L., Zhou, H., Quinn, W. G., & Tully, T. (1994). Induction of a dominant negative CREB transgene specifically blocks long-term memory in *Drosophila*. *Cell*, 79(1), 49–58. [https://doi.org/10.1016/0092-8674\(94\)90399-9](https://doi.org/10.1016/0092-8674(94)90399-9)
- Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14), 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>
- Yudkovsky, N., Logie, C., Hahn, S., & Peterson, C. L. (1999). Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. *Genes & Development*, 13(18), 2369–2374. <https://doi.org/10.1101/gad.13.18.2369>
- Zerbino, D. R., Johnson, N., Juettemann, T., Wilder, S. P., & Flicek, P. (2014). *WiggleTools* : parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. 30(7), 1008–1009. <https://doi.org/10.1093/bioinformatics/btt737>
- Zhang, J., Tanenhaus, A. K., Davis, J. C., Hanlon, B. M., & Yin, J. C. P. (2015). Spatio-temporal in vivo recording of dCREB2 dynamics in *Drosophila* long-term memory processing. *Neurobiology of Learning and Memory*, 118, 80–88. <https://doi.org/10.1016/j.nlm.2014.11.01>

- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). *Open Access Model-based Analysis of ChIP-Seq (MACS)*. 9. <https://doi.org/10.1186/gb-2008-9-9-r137>
- Zhao, B., Sun, J., Zhang, X., Mo, H., Niu, Y., Li, Q., Wang, L., & Zhong, Y. (2019). Long-term memory is formed immediately without the need for protein synthesis-dependent consolidation in *Drosophila*. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-12436-7>
- Zhao, X., Lenek, D., Dag, U., Dickson, B. J., & Keleman, K. (2018). Persistent activity in a recurrent circuit underlies courtship memory in *Drosophila*. *ELife*, 7, 1–16. <https://doi.org/10.7554/eLife.31425>
- Zovkic, I. B. (2021). Epigenetics and memory: an expanded role for chromatin dynamics. *Current Opinion in Neurobiology*, 67, 58–65. <https://doi.org/10.1016/j.conb.2020.08.007>

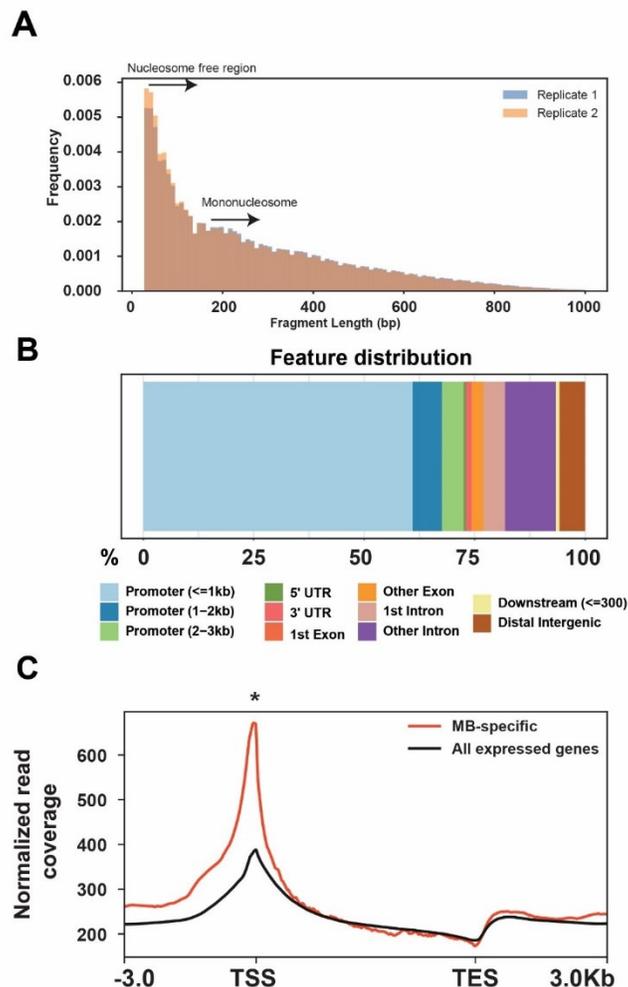
## Appendix A: List of fly stocks used in this project

Controls and genetic tools			
Stock #	Supplier	Genotype	Description
35785	BDSC	$y^l, sc^*, v^l; P\{y^{+t7.7} v^{+t1.8}=VALIUM20-mCherry\}attP2$	$mCherry^{RNAi}$ : short hpRNA UAS-RNAi against $mCherry$ which is not present in the <i>Drosophila</i> genome. TRiP background control for attP2 landing site.
36303	BDSC	$y^l, v^l; P\{y^{+t7.7}=CaryP\}attP2$	TRiP library genetic background control for attP2 landing site
36304	BDSC	$y^l, v^l; P\{y^{+t7.7}=CaryP\}Msp300[attP40]$	TRiP library genetic background control for attP40 landing site. Located in the muscle gene <i>Msp300</i> and may disrupt its function-which is not associated with memory.
n/a	Henry, 2012	$y^l, v^l; P\{UAS\_unc84-2XGFP\}attP2$	$UAS-Unc84-GFP$ : Expressed UNC84-GFP under UAS control.
5137	BDSC	$y^l, v^l; P\{w^{w+mC}=UAS-mCD8::GFP.L\}LL5, P\{UASmCD8::GFP.L\}2$	$UAS-mCD8::GFP$ : Expresses mCD8::GFP under UAS control.
48667	BDSC	$w^{1118} P\{y^{+t7.7} w^{+mC}=GMRI4H06-GAL4\}attP2$	$R14H06-GAL4$ : Expresses GAL4 under the control of a <i>rutabaga</i> (FBgn0003301) enhancer.
25374	BDSC	$y^l, w^*; P\{Act5C-GAL4-w\}E1/CyO$	$Act5C-GAL4$ : Expresses GAL4 ubiquitously under the <i>Act5C</i> (FBgn0000042) promoter.
7019	BDSC	$w^*; P\{w^{+mC}=tubP-GAL80^{ts}\}20; TM2/TM6B, Tb'$	$tubP-GAL80^{ts}$ : Expresses GAL80 <sup>ts</sup> ubiquitously under control of the $\alpha Tub84B$ (FBgn0003884) promoter

RNAi stocks			
Stock #	Supplier	Genotype	Description
63681	BDSC	$y^l, v^l; P\{y^{+7.7} v^{+1.8} = \text{TRiP.HMJ30249}\}attP40/CyO$	<i>CrebB<sup>RNAi</sup></i> : short hpRNA UAS-RNAi against <i>CrebB</i> at attP40 site from the TRiP library. Validation: 100% lethal upon crossing with <i>Act5C-GAL4</i> (n=20)
29376	BDSC	$y^l, v^l; P\{y^{+7.7} v^{+1.8} = \text{TRiP.JF02540}\}attP2$	<i>Hr38<sup>RNAi</sup></i> : long hpRNA UAS-RNAi against <i>Hr38</i> at attP2 site from the TRiP library. Validation: qPCR following <i>Act5C-GAL4</i> knockdown in larvae. 71.75% mRNA knockdown of <i>Hr38</i> compared to mCherry control (p=0.09).
27701	BDSC	$y^l, v^l; P\{y^{+7.7} v^{+1.8} = \text{TRiP.JF02781}\}attP2$	<i>sr<sup>RNAi</sup></i> : long hpRNA UAS-RNAi against <i>sr</i> at attP2 site from the TRiP library. Validation: 100% lethal upon crossing with <i>Act5C-GAL4</i> (n=28)
32503	BDSC	$y^l, sc^*, v^l; P\{y^{+7.7} v^{+1.8} = \text{TRiP.HMS00507}\}attP2$	<i>Bap60<sup>RNAi</sup></i> : short hpRNA UAS-RNAi against <i>Bap60</i> at attP2 site from the TRiP library. Validation: previously published phenotype and qPCR validation (Chubak et al., 2019; Nixon, 2020; Nixon et al., 2019)

## Appendix B: Profiling chromatin accessibility in INTACT-isolated MB nuclei

(A) Fragment size distribution of ATAC-seq libraries generated from INTACT-isolated MB nuclei. Distribution for two biological replicates is shown and is consistent. Peak signal of nucleosome free regions (80-120 bp) and mononucleosomes (~180 bp) is marked. (B) Feature distribution of annotated regulatory regions from significantly accessible peaks and is predominantly TSS-centric, which is expected of ATAC-seq (C) Bandplot profile of normalized read coverage across the gene body for MB-specific genes (n=545) in comparison to all expressed genes (n=6965). MB-specific genes were identified to have significantly greater transcriptional start site accessibility by counting ATAC-seq reads to genes using featureCounts and comparing expression between MB-specific and all-expressed genes (Kruskal-Wallis test,  $p < 0.05$ ). This provides evidence this approach captures MB-specific chromatin accessibility information.



## Appendix C: Code for processing RNA-seq data

RNA-seq reads were processed on Compute Canada servers (StdEnv/2020). First, raw reads were lightly trimmed, and adaptors clipped using Trimmomatic (v0.39)(Bolger et al., 2014). The read quality was then assessed using FastQC (v 0.11.9). Trimmed reads were aligned to the *Drosophila melanogaster* genome (Ensembl release 103, dm6) using STAR (version 2.7.5a) (Aken et al., 2016; Dobin et al., 2013). RNA-seq reads were also aligned to the *C. elegans Unc-84* gene (NC\_003284.9) to validate specificity of MB samples through comparison to WH samples. Uniquely aligned reads with a maximum of four mismatches were counted to genes using featureCounts or HTSeq-count (version 0.7.1) using the default union setting (Anders et al., 2015; Liao et al., 2014).

```
#!/bin/bash
#SBATCH --time=03:00:00
#SBATCH --account=def-jkramer6-ab
#SBATCH --mem-per-cpu=4G
#SBATCH -c 8
#SBATCH -N 1

module load StdEnv/2020
module load trimmomatic/0.39
module load star/2.7.5a
module load samtools/1.11
module load fastqc/0.11.9
module load gcc/9.3.0
module load r-bundle-bioconductor/3.12

var=$(ls *R1.fastq.gz | awk 'BEGIN {FS="_R1.fastq.gz"} {print $1}')
echo $var

gunzip ${var}_R1.fastq.gz
gunzip ${var}_R2.fastq.gz

fastqc ${var}_R1.fastq

java -jar $EBROOTTRIMMOMATIC/trimmomatic-0.39.jar PE ${var}_R1.fastq ${var}_R2.fastq ${var}_forward_paired.fastq ${var}_forward_unpaired.fastq ${var}_reverse_paired.fastq ${var}_reverse_unpaired.fastq ILLUMINACLIP:/home/sjones19/scratch/Adaptors/TruSeq3-PE-2.fa:2:30:10:2:TRUE LEADING:3 TRAILING:3 MINLEN:36

fastqc ${var}_forward_paired.fastq

STAR --runThreadN 8 --genomeDir /home/sjones19/projects/def-jkramer6-ab/genomeDirectory4 --sjdbGTFfile /home/sjones19/projects/def-jkramer6-ab/genomeDirectory4/Drosophila_melanogaster.BDGP6.32.103.UNC84.gtf --readFilesIn ${var}_forward_paired.fastq ${var}_reverse_paired.fastq --outFileNamePrefix ${var} --outReadsUnmapped Fastx --outSAMtype BAM SortedByCoordinate

grep ^@ ${var}Unmapped.out.mate1 | wc -l >> ${var}_stats.out
samtools view -@ 8 -bSq 15 ${var}Aligned.sortedByCoord.out.bam > ${var}_quality.bam
samtools flagstat ${var}_quality.bam | head -n1 |cut -f1 -d + >> ${var}_stats.out
samtools view -@ 8 -h ${var}_quality.bam | grep 'nM:i:0\nM:i:1\nM:i:2\nM:i:3\nM:i:4\n^@' | samtools view -Sb > ${var}.bam
samtools flagstat ${var}.bam | head -n1 |cut -f1 -d + >> ${var}_stats.out
samtools sort -@ 8 -o ${var}.sort.bam ${var}.bam
samtools index ${var}.sort.bam
```

## Appendix D: Code for processing ATAC-seq data

Sequenced ATAC-seq reads were processed on Compute Canada servers (StdEnv/2020). Reads were first lightly trimmed, and adaptors clipped using Trimmomatic (v0.39). Trimmed reads were aligned to the *Drosophila melanogaster* reference genome (Ensembl release 103, dm6) using bowtie2 (v2.4.1) with the settings -X 2000 and -very-sensitive (Aken et al., 2016; Bolger et al., 2014; Langmead & Salzberg, 2012). Reads aligning to multiple loci, the mitochondrial genome, and scaffolds were filtered using samtools view (v1.11) (H. Li et al., 2009). Duplicate reads resulting from PCR amplification were identified using samtools fixmate and removed using samtools markdup. Reads were shifted, +4 bp for the forward strand and -5 bp for the negative strand, to account for the 9-bp duplication created by the DNA repair nick of the Tn5 transpose (Yan et al., 2020).

```
module load StdEnv/2020
module load trimmomatic/0.39
module load bowtie2/2.4.1
module load samtools/1.11
module load fastqc/0.11.9

var=$(ls *_rp.merge_R1.fastq | awk 'BEGIN (FS="_rp.merge_R1.fastq") (print $1)')
echo $var

#fastqc ${var}_R1.fastq

bowtie2 -p 8 -x /home/sjones19/projects/def-jkramer6-ab/sjones19/bowtie2index/genome -1 ${var}.merge_R1.fastq,${var}_rp.merge_R1.fastq -2 ${var}.merge_R2.fastq,${var}_rp.merge_R2.fastq -X 2000 --very-sensitive --no-unal | samtools sort -@ 8 -o ${var}_Aligned.sortedByCoord.out.bam

samtools view -@ 8 -bSq 40 ${var}_Aligned.sortedByCoord.out.bam > ${var}_mapq.bam
samtools sort -@ 8 -o ${var}_mapq.sort.bam ${var}_mapq.bam
samtools index ${var}_mapq.sort.bam
samtools view -@ 8 -b ${var}_mapq.sort.bam 2L 2R 3L 3R 4 X Y > ${var}.bam
samtools sort -@ 8 -n -o ${var}.sort.bam ${var}.bam
samtools fixmate -m ${var}.sort.bam ${var}_fix.bam
samtools sort -@ 8 -o ${var}_fix.sort.bam ${var}_fix.bam
samtools markdup -r ${var}_fix.sort.bam ${var}_rd.bam

#Shifting Reads
#Start by transferring the header to a new sam file
samtools view -@ 8 -H ${var}_rd.bam > ${var}_rd.sam
#Now take the forward reads and shift the start site by 4 bases
samtools view -@ 8 -F 16 ${var}_rd.sam | awk -F'\t' -v OFS='\t' '{$4=$4+4}1' >> ${var}_rd.sam
#Now take the reverse reads and shift the start site by 5 bases
samtools view -@ 8 -f 16 ${var}_rd.sam | awk -F'\t' -v OFS='\t' '{$4=$4-5}1' >> ${var}_rd.sam
#Convert the sam file to bam and then delete the file not needed
samtools view -@ 8 -b ${var}_rd.sam > ${var}_near.bam
samtools sort -@ 8 -n -o ${var}_final.name.bam ${var}_near.bam

samtools view -@ 8 -f 0x02 -b ${var}_final.name.bam > ${var}_nosingles.bam
samtools fixmate -m ${var}_nosingles.bam ${var}_second.fix.bam
samtools sort -@ 8 -o ${var}.final.bam ${var}_second.fix.bam
samtools index ${var}.final.bam

samtools flagstat ${var}_Aligned.sortedByCoord.out.bam | head -n7 | tail -n1 | cut -f1 -d + >> stats_${var}.out
samtools flagstat ${var}_mapq.sort.bam | head -n7 | tail -n1 | cut -f1 -d + >> stats_${var}.out
samtools flagstat ${var}.sort.bam | head -n7 | tail -n1 | cut -f1 -d + >> stats_${var}.out
samtools flagstat ${var}_rd.bam | head -n7 | tail -n1 | cut -f1 -d + >> stats_${var}.out
samtools flagstat ${var}_near.bam | head -n7 | tail -n1 | cut -f1 -d + >> stats_${var}.out
samtools flagstat ${var}.final.bam | head -n7 | tail -n1 | cut -f1 -d + >> stats_${var}.out
```

## Appendix E: MB or WH enriched genes identified using INTACT in adult flies

MB enriched - gene symbol (alphabetical)					
5-HT1A	5-HT1B	a	ab	Ac78C	Act57B
Act87E	Actn	ADPS	Alk	Alr	amon
Antp	Appl	AQP	Asph	asrij	axed
Baldspot	beat-VII	Best1	betaTub56D	bma	bnl
bond	brat	bru2	bru3	bt	btsz
Bx	C3G	Ca-Ma2d	cactin	Cad96Cb	CanA1
capt	CCAP-R	CCHa1-R	CCKLR-17D1	CCKLR-17D3	Ccn
Cda5	Cdk2	CG10019	CG10132	CG10137	CG10483
CG10737	CG10738	CG10916	CG10947	CG11191	CG11200
CG11319	CG11353	CG1136	CG11374	CG11406	CG11550
CG11835	CG11898	CG12038	CG12194	CG12299	CG12531
CG1275	CG12866	CG12910	CG13055	CG13248	CG13284
CG13300	CG13532	CG13618	CG13654	CG13743	CG13921
CG13999	CG14024	CG14186	CG14234	CG14298	CG14434
CG14535	CG14669	CG14692	CG14883	CG14961	CG14964
CG14982	CG1504	CG15097	CG15412	CG15651	CG15765
CG15894	CG1674	CG16868	CG16896	CG17803	CG17816
CG1814	CG18213	CG18304	CG2225	CG2938	CG30116
CG30158	CG30389	CG30419	CG30495	CG31028	CG31140
CG31191	CG31522	CG31523	CG31675	CG31760	CG31817
CG3191	CG32066	CG32085	CG32121	CG32204	CG3253
CG32547	CG32694	CG32700	CG32756	CG32815	CG32944
CG33090	CG33116	CG3339	CG33543	CG3358	CG3408
CG34219	CG34354	CG34357	CG34377	CG34396	CG34402
CG3625	CG3630	CG3823	CG40198	CG40486	CG4115
CG4133	CG4213	CG42238	CG42260	CG4230	CG42322
CG42324	CG42337	CG42339	CG42365	CG4239	CG42390
CG42402	CG42404	CG42588	CG42674	CG43102	CG43331
CG43366	CG43759	CG43861	CG43897	CG44085	CG44247
CG4461	CG44774	CG45076	CG45105	CG45263	CG4562
CG4577	CG46339	CG4660	CG4945	CG5022	CG5023
CG5065	CG5177	CG5191	CG5326	CG5466	CG5535
CG6024	CG6041	CG6044	CG6154	CG6201	CG6428
CG6749	CG6752	CG6765	CG6972	CG7058	CG7130
CG7985	CG8086	CG8248	CG8301	CG8306	CG8412
CG8500	CG8888	CG8910	CG9098	CG9121	CG9123
CG9170	CG9297	CG9313	CG9338	CG9368	CG9492
CG9572	CG9626	CG9760	CG9801	CG9837	cher
ChT	Cht2	Cirl	Cks85A	clumsy	empy
CngA	CngB	Cngl	comm	cora	corn
coro	cos	Cpr51A	Cpr67B	cv-2	cv-c
Cyp4aa1	Cyp6a14	dac	Dbx	Dg	Dgk

Dgp-1	Dh44-R1	Dh44-R2	Dhap-at	Dhc93AB	Dif
DIP-kappa	dl	Dlg5	dlp	dmt99B	DNAPol-eta
DNAPol-iota	DNAPol-zeta	dnc	dock	Dop1R1	Dop1R2
Dop2R	dpr12	dpr14	dpr19	dpr21	dpr5
dpr8	dpr9	Dscam1	Dscam4	dysf	Elk
ey	Fas2	Fatp2	fend	Fife	FipoQ
flr	FMRFaR	fru	Fs(2)Ket	fz	GABA-B-R2
Gadd45	GC	Gfr1	gkt	GlcAT-P	GluRIA
GluRIB	GlyT	gprs	Grik	grk	grp
Gsc	GstD11	GstT4	hec	Hr38	Hr51
Hs3st-B	Hsc70-3	Hsp26	Hsp70Bb	hzg	Imp
IntS12	Ir68a	Ir93a	Irbp	jdp	jeb
JhI-21	jus	Kah	KaiR1D	kat-60L1	kay
kek1	kel	kni	knrl	Ku80	kuz
l(2)01289	l(2)k01209	lbm	LBR	Ldh	Lgr1
Lgr3	Lrp4	Lrrk	LUBEL	LysRS	M6
mAChR-A	mAChR-B	Magi	mamo	Marf1	Mef2
meng	metro	MetRS	MFS3	Mhc	Mical
mip120	Mkp3	Mlc1	Mlp60A	mmd	Mp20
Mpcp1	Msp300	mspo	mtd	mthl8	mtt
mub	Myo81F	na	nAChRalpha1	nAChRalpha2	nAChRalpha5
nAChRalpha6	nAChRbeta2	Ndg	Nep1	net	Neurochondrin
Nlg1	Nlg3	Nmdar1	nompB	nompC	nord
Nrk	nvy	Oamb	Octbeta1R	olf413	Oli
ome	Orect2	Oseg2	Oseg6	otk	Patsas
Pde8	pdm2	Pgant9	PH4alphaNE1	pHCl-1	phyl
PICK1	Pka-C1	Pka-R1	Pka-R2	Pkc53E	Pkcdelta
Plc21C	plh	plx	POSH	Ppcs	prage
Prm	promL	prt	PVRAP	pxb	Pxn
Pzl	qin	Rab26	Rab27	rad	Rbp
rdgBbeta	Rdl	rempA	Rep	Rgk1	Rgk2
Rgk3	Rgl	rgn	rho-5	RhoGAP100F	Rim
rk	robo1	rod	Ror	Rsph3	rt
Rtca	rut	RyR	Sap130	Sc2	sca
sced	Scgbeta	Scgdelta	Sema2b	SERCA	Shab
Shaw	Shawl	sick	side-II	side-III	side-VI
side-VII	SIFaR	Sil1	Sln	SLO2	sls
smo	smog	Smyd4-1	SmydA-9	Snap24	snoRNA:U3:9B
sNPF	snRNA:7SK	SP1173	spab	spas	Spg7
spidey	spirit	sqa	sr	Stacl	Strn-Mlck
Stt3A	Syt12	Syt14	Syalpha	Sybeta	Task6
Task7	tbc	Ten-a	Tg	Tm2	Tmc
tn	toc	Toll-7	Tomosyn	toy	TpnC25D
TpnC73F	trh	TrpA1	Trpm	trv	Trxr-2
Tsen34	Tsp42Eg	Tsp42Eh	Tsp96F	TTLL15	TTLL3A

TwdlG	tyn	Ude	unc-13-4A	unc-5	Unc-89
unc79	unc80	up	Urod	ValRS	VepD
vg	viaf	Vsx1	wat	Wnt6	wry
wupA	Zasp66	zfh2	zormin	zye	

WH enriched - gene symbol (alphabetical)					
26-29-p	Acbp1	Acbp2	Acer	Achl	acj6
Acox57D-p	Acyp2	AdenoK	Adgf-D	Adi1	AdipoR
Adk2	Agpat4	Ahcy	AkhR	Akr1B	Aldh7A1
Alp4	alpha-Est1	alpha-Est2	alpha-Est5	alpha-Est7	alpha-Est8
Amy-d	Amy-p	ana	Ance	Ance-4	Ance-5
antdh	AOX1	ap	apolpp	Apoltp	Aprt
Arc42	Arg1	Arr1	AstA	AstA-R1	AstC-R1
Atg8a	atk	ATPsynCF6	ATPsynD	ATPsyndelta	ATPsynE
ATPsynF	ATPsynG	AttA	AttB	AttC	awd
axo	bd1	ben	beta-Man	betaTub97EF	bgm
bic	BomBc1	BomBc2	BomBc3	BomS1	BomS2
BomS3	BomS5	BomS6	BomT2	BomT3	boss
bsh	bw	Caf1-180	CAH1	CAH2	CAH5
Cam	CarT	Cat	cathD	CCHa2	Ccm3
CDase	Cds	CecA1	ced-6	CG10031	CG10166
CG10184	CG10237	CG10343	CG10345	CG10348	CG10352
CG10361	CG10433	CG10479	CG10513	CG10516	CG10553
CG10560	CG10621	CG10660	CG10680	CG10799	CG10863
CG1092	CG10932	CG11089	CG11211	CG11236	CG11267
CG11294	CG11313	CG11368	CG11400	CG11407	CG11529
CG11594	CG11752	CG11841	CG11842	CG11899	CG11951
CG12091	CG1213	CG12171	CG12239	CG12338	CG12384
CG12512	CG12811	CG13077	CG13082	CG13086	CG13220
CG13305	CG13315	CG13360	CG13364	CG13397	CG13428
CG13551	CG13585	CG13631	CG13704	CG13707	CG13722
CG13794	CG13795	CG13833	CG13912	CG14259	CG14274
CG14275	CG14400	CG14407	CG1441	CG14591	CG14615
CG1468	CG14688	CG14715	CG14812	CG14818	CG14977
CG15019	CG15096	CG15098	CG15117	CG15118	CG15186
CG15201	CG15203	CG15347	CG15414	CG1544	CG15478
CG1552	CG15553	CG15739	CG15771	CG16704	CG16713
CG1673	CG16743	CG16756	CG16758	CG16772	CG16820
CG16898	CG16926	CG16965	CG16978	CG16986	CG17005
CG17108	CG17167	CG17193	CG17224	CG17278	CG17333
CG17378	CG17572	CG17597	CG17739	CG1774	CG17896
CG1791	CG17928	CG18003	CG18067	CG18081	CG18135
CG18249	CG18302	CG18343	CG18508	CG18547	CG18622
CG18815	CG1889	CG2004	CG2016	CG2082	CG2145

CG2233	CG2493	CG2681	CG2736	CG2767	CG2811
CG30002	CG30026	CG30033	CG30046	CG30203	CG30431
CG3091	CG31036	CG31038	CG31075	CG31102	CG31126
CG31207	CG31248	CG31313	CG31326	CG31436	CG31445
CG31548	CG31636	CG31663	CG31673	CG3168	CG31705
CG31706	CG31777	CG31778	CG31937	CG32039	CG32091
CG32191	CG32267	CG32276	CG32278	CG32373	CG32407
CG32444	CG3246	CG32687	CG32695	CG3270	CG32726
CG33080	CG33110	CG33120	CG33296	CG33493	CG33494
CG33639	CG3376	CG33926	CG34117	CG34134	CG34165
CG34166	CG3420	CG34200	CG34227	CG34228	CG34330
CG34331	CG3505	CG3534	CG3566	CG3597	CG3609
CG3663	CG3699	CG3700	CG3829	CG3902	CG3999
CG4000	CG40472	CG40485	CG41128	CG42329	CG42331
CG4250	CG42846	CG42876	CG4306	CG43103	CG43340
CG4364	CG4390	CG44014	CG4408	CG4409	CG44242
CG4462	CG4572	CG4593	CG4598	CG4607	CG46448
CG4646	CG4716	CG4752	CG4757	CG4842	CG4847
CG4872	CG5009	CG5059	CG5110	CG5162	CG5171
CG5254	CG5273	CG5287	CG5321	CG5377	CG5390
CG5404	CG5577	CG5590	CG5707	CG5773	CG5793
CG5819	CG5835	CG5849	CG5862	CG5895	CG5945
CG5955	CG5958	CG5966	CG5973	CG6028	CG6067
CG6126	CG6180	CG6191	CG6218	CG6287	CG6357
CG6409	CG6421	CG6426	CG6429	CG6435	CG6465
CG6503	CG6523	CG6565	CG6638	CG6656	CG6723
CG6770	CG6805	CG6830	CG6834	CG6870	CG6910
CG6983	CG7016	CG7079	CG7084	CG7149	CG7220
CG7255	CG7299	CG7322	CG7365	CG7461	CG7470
CG7484	CG7509	CG7560	CG7630	CG7675	CG7724
CG7800	CG7829	CG7834	CG7966	CG7997	CG8036
CG8051	CG8066	CG8080	CG8128	CG8132	CG8199
CG8206	CG8343	CG8360	CG8369	CG8419	CG8468
CG8539	CG8586	CG8654	CG8665	CG8738	CG8916
CG8993	CG9034	CG9231	CG9312	CG9331	CG9391
CG9396	CG9427	CG9436	CG9449	CG9451	CG9471
CG9485	CG9497	CG9498	CG9527	CG9631	CG9649
CG9657	CG9686	CG9689	CG9691	CG9917	CG9922
CG9928	CG9953	Chd64	chp	CIA30	Cisd2
cl	CIC-a	Clk	CNBP	Cnx99A	Col4a1
colt	COX4	COX5A	COX5B	COX6B	COX7A
COX7C	COX8	Cp1	Cpn	Cpr	Cpr100A
Cpr49Ae	Cpr72Ec	Cpr76Bc	CRAT	CREG	Crg-1
cry	Crys	Csas	ct	CtrlB	CtsB1
Culd	cv-d	Cyp1	Cyp12a4	Cyp12c1	Cyp18a1

Cyp28a5	Cyp28d1	Cyp305a1	Cyp309a1	Cyp309a2	Cyp311a1
Cyp4d21	Cyp4d8	Cyp4e2	Cyp4e3	Cyp4g15	Cyp4s3
Cyp6a21	Cyp6a22	Cyp6a8	Cyp6d4	Cyp6g1	Cyp6w1
Cyp9b1	Cyp9b2	cype	Cyt-b5	Cyt-b5-r	D
D2hgdh	Dak1	DCX-EMAP	Dera	Desat1	Dhc98D
DIP-eta	disco	dj-1beta	DnaJ-1	DNaseII	dob
DptA	DptB	Drep3	Drgx	Dro	Drs
Drsl4	Drsl5	e	Eaat1	Echs1	eEF1alpha1
eEF1beta	eEF1delta	eEF1gamma	eEF5	eIF3k	Eip55E
Ekar	Elal	EloB	EloC	EMC6	erm
Est-6	et	Etf-QO	Ets65A	euc	exp
eyg	eyS	Faa	FASN3	fat-spondin	fbp
Fbx14	fd59A	Fdh	Fdx2	Fer1HCH	Fis1
fit	Fkbp12	Fkbp14	flw	Fmo-1	Fmo-2
fon	frma	fusl	Gale	Galk	galla-2
gammaSnap1	Gart	Gat	Gba1b	Gbeta76C	Gbp1
Gbp2	Gbp3	Gbs-70E	Gclm	Gdh	Gel
Ggamma1	Ggamma30A	GILT2	Gip	Gk2	gl
GLaz	Gld	Glo1	glob1	Glt	GNBP-like3
GNBP1	GNBP2	GNBP3	Gnmt	Gnpnat	gol
Got1	Got2	Grd	Grx1	Gs2	GstE1
GstE12	GstE14	GstE2	GstE3	GstE7	GstE8
GstE9	GstS1	GstZ1	Hayan	hbn	HDAC6
Hdc	Hexo1	Hexo2	Hf	hgo	HINT1
His3.3A	His4r	Hml	Hmx	Hn	homer
HP4	Hpd	IBIN	Idgf2	Idgf3	Idgf4
Idgf5	Idh	IM14	IM33	IM4	ImpE1
ImpL2	inaC	inaD	Inx2	Ipk1	Ipp
Irc	Iris	Irk3	ITP	Jabba	Jafrac1
janA	Kaz1-ORFB	KFase	Kyat	l(1)10Bb	l(3)77CDF
LanA	LanB2	Ldsdh1	lectin-28C	levy	Lim3
Lip2	Listericin	LManI	LManII	lovit	lqf
Lsd-1	Lsd-2	Lsp1alpha	Lsp1beta	Lsp2	Lst
lz	maf-S	Mal-A5	Marc	Mccc2	Mdh1
mdy	Men	Mes2	MFS14	MFS9	mge
Mgstl	mino	Mip	Mob2	Mob3	Mpc1
Mppe	mRpL24	mRpL27	mRpL33	mRpL42	mRpL51
mRpL55	mRpS14	mRpS17	mRpS33	mRpS35	Ms
msi	MsR2	MsrA	Mtap	Mthfs	MtnA
Mtp	Mtpalpha	Mtpbeta	Muc14A	mv	Myo28B1
Nap1	ND-13A	ND-13B	ND-15	ND-20	ND-ACP
ND-AGGG	ND-ASHI	ND-B12	ND-B14	ND-B14.5A	ND-B14.5B
ND-B14.7	ND-B17	ND-B17.2	ND-B18	ND-B22	ND-MLRQ
ND-MWFE	ND-PDSW	ND-SGDH	Neb-cGP	Nedd8	Nep7
Nep112	Nep19	NimB2	NimB4	NimB5	NimC1

ninaB	ninaC	ninaE	ninaG	Nmdmc	norpA
Npc2a	Npc2g	Npc2h	NPF	Nph	Nplp2
Nplp3	nrv2	NT5E-2	NtR	Nurf-38	Oat
Oatp58Dc	Obp18a	Obp19b	Obp19d	Obp44a	Obp56g
Obp57a	Obp57b	Obp57c	Obp99b	Obp99c	oc
Oda	Odc1	ogre	opa	ort	ox
Paics	path	PCB	Pcd	Pdf	Pdh
Pdk	Pdxk	peb	Pfas	Pglym78	Pgm1
PGRP-LA	PGRP-SA	PGRP-SB1	PGRP-SC2	phu	pinta
Pis	Pisd	Pkn	PLCXD	ple	pn
pnr	pnt	Pomp	Pph13	ppl	Ppn
Ppt2	pr	Prat2	prom	Prosbeta3	Prx2540-2
psh	pst	Pu	pug	Qsox1	Rab32
Rab5	Rab7	rdgA	rdhB	Reg-2	regucalcin
REPTOR-BP	retn	rgr	Rh2	Rh3	Rh4
Rh5	Rh6	RhoGAP18B	roh	Rpb12	Rpi
RpII18	RpL10	RpL11	RpL12	RpL13A	RpL14
RpL15	RpL17	RpL18	RpL18A	RpL19	RpL21
RpL23	RpL24	RpL24-like	RpL26	RpL27	RpL27A
RpL28	RpL29	RpL30	RpL31	RpL32	RpL34a
RpL34b	RpL35	RpL36	RpL36A	RpL37a	RpL38
RpL39	RpL40	RpL5	RpL6	RpL7	RpL9
RpLP0-like	RpLP1	RpLP2	RpS10b	RpS11	RpS13
RpS14a	RpS14b	RpS15	RpS15Aa	RpS15Ab	RpS17
RpS18	RpS19a	RpS20	RpS21	RpS23	RpS25
RpS26	RpS27	RpS27A	RpS28b	RpS29	RpS3
RpS30	RpS3A	RpS5a	RpS7	RpS8	RpS9
Rpt6	rtp	ry	santa-maria	Sap-r	Sar1
Sardh	Sccpdh2	SCOT	Scp2	scro	Scsalphal
ScsbetaA	scu	sds22	se	Sec61beta	Sec61gamma
SelG	SelT	Sem1	Ser	Ser7	Sfxn1-3
shg	Shmt	SiaT	sim	Sirup	SkpA
SLC5A11	smt3	SmydA-8	sn	sni	SNRPG
Snx3	so	Sod1	Sod3	Sodh-1	sosie
Sox21b	SoxN	SPARC	Spat	SPE	sPLA2
Spn28Dc	Spn42Dd	Spn88Eb	Sr-CI	srp	Srp19
Srp9	Srr	Ssb-c31a	ssp6	St1	sta
su(r)	sug	sun	swi2	Synd	Syx8
t	Tab2	Taf10	Taldo	Tctp	Tep2
Tep4	teq	TfAP-2	TfIIA-S	Tig	tj
Tk	to	tobi	toe	Tom7	TotA
TotC	TotM	Tps1	trp	trpl	Tsfl
tsl	Tsp5D	Ubc10	Uev1A	Ufc1	Ugt301D1
Ugt35B1	Ugt35C1	Ugt36E1	Ugt37B1	Ugt49B1	Ugt50B3
UK114	UQCR-11	UQCR-6.4	UQCR-Q	Usp8	v

Vamp7	VGlut	Vinc	vkg	Vti1b	wal
wdp	wrapper	wun2	yellow-e	yellow-f2	yellow-h
yip2	Yp3	zyd			

## Appendix F: TIGs identified in both the WH and MB, WH-only or MB-only

TIGs identified in both the WH and MB – gene symbol (alphabetical)					
AANAT1	AcCoAS	Achl	Act79B	Act87E	Actn
AdSS	Agpat2	Aldh-III	alpha-Est10	alpha-Est8	Ance-3
AnxB10	AnxB11	AnxB9	aop	Apoltp	Arf79F
Argk	Asph	AstA	Atf3	atk	ATPsynF
b	bark	bel	beta4GalNAcTA	betaTub97EF	bgm
bic	blot	bon	bur	cact	Cdc27
Cdc7	CG10026	CG10205	CG10237	CG10341	CG10345
CG10348	CG10512	CG10527	CG10550	CG10560	CG10621
CG10737	CG10947	CG11050	CG11241	CG11368	CG11529
CG11899	CG12512	CG12814	CG13096	CG13185	CG13315
CG13360	CG13428	CG13784	CG13794	CG13912	CG14186
CG14207	CG14277	CG14401	CG14439	CG14661	CG14687
CG14688	CG14694	CG14894	CG15021	CG15096	CG15544
CG15673	CG1572	CG1628	CG1648	CG1674	CG16743
CG16820	CG17111	CG17193	CG17221	CG17278	CG18622
CG2016	CG2201	CG2865	CG30172	CG3036	CG31121
CG31126	CG31324	CG32032	CG32091	CG32276	CG32369
CG32425	CG32694	CG33110	CG33296	CG3348	CG33970
CG3597	CG3630	CG3764	CG3842	CG40006	CG4199
CG42390	CG42672	CG43324	CG43658	CG43693	CG4374
CG44242	CG44247	CG4461	CG4554	CG4577	CG4733
CG4797	CG4968	CG5009	CG5023	CG5033	CG5059
CG5065	CG5080	CG5104	CG5151	CG5177	CG5191
CG5346	CG5377	CG5597	CG5618	CG5707	CG5773
CG5849	CG5909	CG5945	CG5955	CG5958	CG5973
CG6142	CG6145	CG6175	CG6701	CG6910	CG6972
CG7365	CG7530	CG7720	CG7920	CG8008	CG8051
CG8157	CG8306	CG8369	CG8654	CG8665	CG8745
CG8939	CG9005	CG9034	CG9253	CG9399	CG9436
CG9451	CG9498	CG9674	CG9717	CG9812	CG9815
CG9837	cher	ClC-c	CLIP-190	Clk	coro
COX5B	COX8	Cpr	Cpr67B	CrebA	Crg-1
CROT	cwo	Cyp4d21	D2hgdh	daw	Dgk
Diap1	dl	Dll	DOR	Dp1	Drat
drm	Droj2	dsb	dsx	Duox	EbpIII
edl	egr	eIF1A	Elal	ELOVL	Eno
Ent2	Esp	FASN1	Fatp3	fbl	Fer1HCH
FipoQ	fit	Fmo-2	fng	fry	Gclc
Gdh	GILT1	Gk2	GlyP	Gpdh1	Gpo1
GstD1	gukh	h	Hacd1	Hex-A	Hmgs
Hr38	Hsc70-4	Idh	imd	ImpL2	insec
Irk1	Irp-1B	ITP	Jafrac1	jeb	Jra

Kah	Kaz1-ORFB	kdn	knrl	kon	Kr-h1
Kr-h2	krz	l(2)01289	l(3)77CDF	Lac	LanB2
lbk	Lgr1	LKRSDH	loco	Lrt	Lsd-1
LUBEL	mAcon1	Max	Mcol	Mctp	Men
meng	Mes2	MESR3	MFS14	MFS9	Mip
Mlc2	MME1	Mmp1	Mocs1	mod	Moe
mol	Mp20	Mrp4	msh	MsrA	Mtpalpha
muc	Myo31DF	ND-49	ND-B12	Nep2	Nep7
Neurochondrin	Nha1	NimC1	NKCC	Nlp	Nop60B
Nopp140	Ns4	Obp49a	Obp56d	Obp56g	out
oys	pain	PCB	Pdh	Pdk1	Pect
per	pes	Pgd	Phk-3	phu	Picot
Pino	Pisd	pit	ple	plh	Pli
plx	Pmp70	pot	Ppcs	ppl	Ppn
PRAS40	Prm	pst	Pu	puc	puml
Pura	Pvr	PyK	qm	r	raw
rictor	Root	RpI1	RpL29	RpL31	RpS15
RpS17	RpS19a	RpS9	ry	S6k	Sam-S
Sardh	sbm	SCaMC	Scp1	Sec63	SelG
SERCA	SerT	shn	Sidpn	Sik3	slf
slow	SmydA-9	snu	Sobp	Sod3	Sp1
spin	Sply	Spn42Dd	spz	sr	srl
sty	su(r)	subdued	sug	svr	Tep3
Thor	Tig	tim	tn	tnc	trc
Tre1	Treh	Trp1	Trxr-1	Tsp42Ee	Tsp42Ef
Tsp42Eg	Tsp42Ei	Ttd14	Tudor-SN	Ude	Ugt302C1
UQCR-C1	Usp1	Vha100-2	Vha36-1	Vinc	vir-1
Vps13	Vps60	wat	wdp	Wnt6	Zasp66
zfh1	ZnT63C	ZnT77C			

TIGs identified only in the WH – gene symbol (alphabetical)					
14-3-3epsilon	14-3-3zeta	Abi	Acbp1	ACC	Acn
Adi1	Adk2	Aef1	AGO1	Agpat1	Agpat4
akirin	Alg1	ALiX	alph	Amun	AOX1
AOX3	Aps	AQP	aqz	Arf51F	ari-1
Arl1	Arl8	AstC	Atg1	Atg17	Atg6
Atg8a	Atpalpha	ATPCL	ATPsyndelta	ATPsynE	Atx2
Bacc	bai	bc10	boca	bocks	bol
BomS1	BomS2	BomS3	Bsg	bsk	BTBD9
c12.2	CAH2	CAH3	Cam	CanA-14F	CanB
CARPA	casp	cathD	Cbl	Cbp20	Ccm3
CCT7	Cdc42	cdm	cg	CG10209	CG10333
CG10340	CG10352	CG10395	CG10433	CG10470	CG10508
CG10576	CG10600	CG10680	CG10960	CG11000	CG11076

CG11180	CG11200	CG11211	CG11267	CG11317	CG11334
CG11378	CG11417	CG11425	CG1143	CG11583	CG1161
CG11791	CG11811	CG11854	CG11857	CG11927	CG12071
CG12093	CG12173	CG12213	CG12239	CG12301	CG12325
CG12355	CG1236	CG12605	CG12608	CG12811	CG12866
CG13024	CG1316	CG13220	CG13364	CG13367	CG13585
CG13606	CG13625	CG13631	CG13722	CG13751	CG13887
CG13994	CG14024	CG14132	CG14184	CG14274	CG14275
CG14341	CG1441	CG14435	CG14478	CG14591	CG1468
CG14715	CG14830	CG14892	CG14989	CG15019	CG15201
CG15202	CG15279	CG15312	CG1532	CG15602	CG15715
CG15765	CG15771	CG1607	CG16772	CG16953	CG1703
CG17124	CG17202	CG17271	CG17321	CG17385	CG17454
CG17574	CG17734	CG17778	CG1785	CG1789	CG17919
CG18067	CG18178	CG18428	CG18815	CG1910	CG2065
CG2162	CG2199	CG2608	CG2662	CG2681	CG2811
CG2918	CG2972	CG30026	CG30033	CG3008	CG30109
CG30151	CG30285	CG30389	CG30423	CG3061	CG31248
CG31694	CG31729	CG31777	CG31886	CG3198	CG31997
CG32278	CG32344	CG32576	CG32647	CG32726	CG33181
CG3335	CG34134	CG34200	CG34228	CG34331	CG3760
CG3829	CG3907	CG4115	CG4213	CG42240	CG42258
CG4239	CG42795	CG43066	CG4328	CG4364	CG44838
CG45050	CG45076	CG4511	CG4612	CG46338	CG4806
CG4829	CG4945	CG5162	CG5273	CG5316	CG5522
CG5567	CG5645	CG5800	CG6115	CG6201	CG6236
CG6329	CG6330	CG6357	CG6429	CG6656	CG6665
CG6712	CG6805	CG6843	CG6870	CG6912	CG6983
CG7016	CG7115	CG7120	CG7130	CG7168	CG7203
CG7220	CG7272	CG7275	CG7339	CG7378	CG7484
CG7509	CG7546	CG7560	CG7630	CG7724	CG7772
CG7778	CG7800	CG7872	CG7878	CG7990	CG8036
CG8149	CG8188	CG8206	CG8303	CG8360	CG8389
CG8399	CG8405	CG8414	CG8485	CG8611	CG8668
CG8677	CG8814	CG8839	CG8910	CG8963	CG9004
CG9065	CG9132	CG9297	CG9336	CG9485	CG9572
CG9586	CG9705	CG9706	CG9813	CG9849	CG9911
CG9922	Chc	Chchd2	Chd64	chic	Cht11
CIA30	Cka	CkIalpha	CkIIbeta	CNBP	Cnot4
Cnx99A	coil	Col4a1	comm2	corto	Cow
COX5A	COX6B	COX7C	Cpn	Cpr47Ee	Cpr49Ae
CRIF	Crtc	CSN6	ctp	CTPsyn	cv-2
cv-c	cv-d	CycT	Cyp1	Cyp4e1	Cyp4e2
Cyp4e3	Cyp6w1	cype	Cyt-b5	Cyt-c-p	D1
Dad1	dbe	DCP2	Deaf1	Dh44	Dhit

Dic1	Dip-C	Dlc90F	Dlg5	Dlic	dmpd
dmrt99B	DopEcR	Dp	Dph5	dpr20	Drep1
dtm	dysc	e(r)	Edc3	eEF1alpha2	eEF1delta
eff	egl	eIF1	eIF2beta	eIF3g1	eIF3h
eIF3j	eIF3k	eIF4E1	eIF4G1	eIF4H1	eIF5B
EloB	Elp2	EMC4	EMC5	EMC6	EMC7
emp	endos	ens	eRF3	ergic53	Faf2
Fatp2	fax	FBXO11	fd102C	Fdx2	Fer2LCH
Fib	Fip1	Fkbp12	Fkbp14	Fkbp39	Flo2
flw	Fmr1	fon	FRG1	frma	Frq1
Frq2	Gadd45	Galphas	gammaSnap1	Gasp	gb
Gdap2	Gfat1	Ggamma30A	ghi	GIIIspla2	glec
Glut4EF	GNBP2	Gnmt	gol	Gp210	Grik
Gs2	GstD9	GstE3	GstO3	GV1	Hdc
HIP	His3.3A	His3.3B	His4r	Hk	HmgZ
holn1	homer	hook	Hr3	Hrb27C	Hrb87F
Hrb98DE	hrg	Hsc70-1	Hsf	hwt	Hydr2
Idgf3	Idgf4	Idh3a	Idi	lh	Imp
inaD	ine	Ing5	Inos	IscU	JTBR
Jwa	Kap-alpha3	KdelR	kek1	ken	Klp98A
kra	kraken	l(1)G0020	l(1)G0289	l(1)G0320	l(2)09851
l(2)k01209	Lamp1	Larp4B	laza	lbn	Letm1
levy	Lk6	loj	lola	lolal	Lsd-2
Lsm12a	Lst	ltl	maf-S	mahe	mamo
Manf	Map205	Mapmodulin	Mat1	MBD-like	Mdr49
MED22	MED6	Meics	meso18E	metl	mge
Mgstl	Mhc	MICAL-like	miple1	Mkp3	Mlc1
Mlf	Mlp60A	Mnn1	Mnt	Mo25	Mob2
MrgBP	mRpL18	mRpL19	mRpL24	mRpL27	mRpL30
mRpL33	mRpL47	mRpL9	mRpS11	mRpS14	mRpS26
mRpS33	mrva	Ms	MSBP	Mtap	MTF-1
MtnA	Mtp	mts	mura	Mvl	Naa30A
Nap1	Naus	Nca	nclb	ncm	ND-ASHI
ND-B14	ND-B17	ND-B18	Ndf	Neb-cGP	Nelf-A
nero	nes	NimB2	ninaE	nmd	Nmd3
Nnp-1	NO66	Nop56	norpA	Nph	Nplp2
Nplp3	nrv3	nSyb	Ntf-2	NUCB1	nudE
Obp18a	Obp19b	Obp56h	Obp99c	oc	Oda
Ostgamma	ox	p24-1	pAbp	PAPLA1	path
Pc	Pcmt	PDCD-5	Pdi	Pdp1	Pepck2
Pex5	Pfdn1	Pfdn2	Pfdn4	Pfdn5	Pglym78
Pits	poly	Pop1	porin	Pp1-87B	pr
prage	Proc	Prx2540-2	ps	Psc	Ptpmeg2
pyr	qkr58E-1	qvr	Rab10	Rab11	Rab14
Rab2	Rab3	Rab5	Rab6	Rab7	RabX1

RabX4	Rac1	RagC-D	Ran	RanBPM	Rap1
Rap21	RasGAP1	Rb97D	Rbm13	Rbp1-like	Rbp6
Rdl	rdog	ReepB	REG	Reg-5	Reph
Rh6	Rheb	Rho1	RhoGAP15B	RhoGAP5A	RhoGAP68F
Rilpl	Rim	Rip11	Rlb1	Rm62	robl
roh	Rok	roq	Rpb12	RpII18	RpL12
RpL14	RpL24-like	RpL28	RpL6	RpL9	RpLP0-like
Rpn12	Rpn9	RpS11	RpS13	RpS20	RpS21
RpS23	RpS27	RpS28b	RpS29	RpS7	Rpt6
Rsl	Rsf1	rudhira	salm	SamDC	san
Sap47	Sar1	Sas10	sau	Sbp2	Scamp
Sccpdh1	SCCRO4	scf	schlank	ScpX	scrt
scyl	sd	SdhC	sds22	Sec22	Sec61alpha
Sec61beta	Sec61gamma	Sem1	Ser	sev	Sf3b6
sgg	Sgt	Sh	shep	sip3	Sirup
SK	SkpA	Slik	smt3	Snp	snRNP-U1-C
SNRPG	Snx3	SoxN	Sp7	spab	Spase22-23
Spg7	spirit	sqd	Srp14	Srp19	Srp9
SRPK	Ssdp	SsRbeta	ssx	Stam	stau
Sting	STUB1	Su(dx)	Sugb	sun	Surf4
sws	Synd	Syx5	Taf10	Tango5	Tapdelta
tau	tbc	Tep4	TfIIA-S	TfIIFalpha	TfIIS
tgo	Tlk	Tm2	TM4SF	TM9SF4	Tmc
Tmem63	TMS1	tna	tok	Tom40	Tom7
Top1	TpnC41C	tral	Trf2	Trf4-1	Trx-2
Tsp97E	Tspo	tsr	tyf	U2af50	Ubc10
Ubc4	Ubc6	UbcE2H	UbcE2M	Ubi-p63E	UBL3
Ubqn	Ufc1	unc-4	up	UQCR-11	UQCR-6.4
Usp32	Usp8	Vap33	VepD	Vha13	Vha16-1
Vha44	vig2	vkg	Vmat	Vps26	Vps29
vri	Vrp1	vsg	Vsx1	vvl	wey
wdb	wds	Wnk	wupA	Xrp1	yellow-c
Yeti	Ykt6	zetaCOP	Zip89B	zye	

TIGs identified only in the MB – gene symbol (alphabetical)					
5-HT1A	a	Ac78C	Acs1	AdamTS-B	Adgf-A
ADPS	AkhR	Akr1B	ap	aralar1	Arpc2
Art3	ash2	ATPsynbeta	ATPsyngamma	axo	Baldspot
baz	bd1	blw	btl	btsz	by
Cad86C	CAP	Ccdc85	Cen	CG10055	CG10126
CG10311	CG10513	CG10553	CG10939	CG11073	CG11313
CG11319	CG11550	CG11980	CG12065	CG12268	CG12290
CG12344	CG12531	CG1265	CG12746	CG13124	CG13284
CG13579	CG1358	CG13743	CG14220	CG14407	CG14762
CG15186	CG1552	CG15528	CG15890	CG1640	CG1673
CG17549	CG17572	CG17646	CG17816	CG17834	CG18135
CG1814	CG2004	CG2082	CG2841	CG2921	CG30069
CG30158	CG31036	CG31038	CG31103	CG31436	CG31522
CG31523	CG31663	CG31689	CG31760	CG32204	CG32547
CG3339	CG33521	CG33978	CG3625	CG3823	CG41520
CG42319	CG42339	CG42534	CG42673	CG43078	CG44085
CG44245	CG44325	CG4502	CG45105	CG46385	CG4660
CG4928	CG4991	CG5028	CG5214	CG5326	CG5382
CG5455	CG5599	CG5687	CG5721	CG5787	CG5867
CG6006	CG6040	CG6123	CG6126	CG6225	CG6231
CG6356	CG6847	CG7133	CG7299	CG7322	CG7888
CG7900	CG7997	CG8034	CG8042	CG8180	CG8289
CG8500	CG8520	CG8547	CG8768	CG9338	CG9449
CG9626	chb	CHKov1	Chmp1	Clc-a	ClpP
cnc	CngA	cnk	Coq8	CoRest	Cpr72Ec
CPT2	crb	Cyp12a5	Cyp28d1	Cyp309a1	Cyp4aa1
Cyp9h1	CysRS	dac	DAT	Dcr-2	DCTN5-p25
DCX-EMAP	Dhc93AB	DhpD	dia	DIP-eta	Dmtn
dnr1	dom	Dop1R2	dpr4	Drip	e
Eaat1	Edem1	eIF3b	eIF3f1	Elk	emb
Ets65A	ey	FER	Fhos	firl	fkf
foxo	Fum1	fw	GABA-B-R3	Gapdh1	Gga
Gli	grass	Grip	grk	ham	HDAC1
hec	Hil	how	Hr51	Hsp68	htl
hui	if	Irk2	Jabba	jdp	Keap1
kn	kni	Ktl	l(3)05822	lbl	Lgr3
Loxl2	Lpin	lwr	Mat89Ba	Mct1	metro
Mi-2	Mlc-c	Mrtf	Mtpbeta	NAAT1	nahoda
Naprt	Nc73EF	Ncc69	ND-75	ND-B14.5A	ND-PDSW
Ndae1	Nedd4	nompC	Npc2h	NPF	Npl4
nrv1	nrv2	Ns2	Nsf2	NtR	Nuak1
Nup54	navy	Octbeta3R	Odc1	ome	ort
Pax	pbl	Pde11	Pex2	Pfk	Pgk
pio	Pka-C1	Pka-R1	Pkcdelta	Plc21C	Proc-R

Pxn	r-l	Rat1	rau	Rbp	Rbsn-5
Rcd2	rdgB	Rgk1	Rgk2	Rgk3	Rgl
rgr	Rh50	robo3	RpA-70	RpL17	RtGEF
rut	Rx	Sarm	scaf	SdhA	SdhB
sdt	Sec5	sls	SmydA-7	Snap24	snoRNA:Psi28S-3342
sNPF	snRNA:U2:34ABa	sosie	Sox15	SP1029	spg
sqa	SREBP	ss	stac	Strn-Mlck	Svil
svp	Syng1	Syx17	Tab2	tacc	Taf12
Tak1	Ten-m	Tg	Tina-1	TM9SF2	Traf4
trpl	tsl	Tsp42Eh	tup	uif	unc-104
Unc-89	UQCR-C2	uzip	Vha14-1	Vha36-3	Vha68-2
VhaSFD	wmd	Zasp52			